

AN INVERSE THEOREM FOR THE GOWERS  $U^{s+1}[N]$ -NORM

BEN GREEN, TERENCE TAO, AND TAMAR ZIEGLER

ABSTRACT. We prove the *inverse conjecture for the Gowers  $U^{s+1}[N]$ -norm* for all  $s \geq 1$ ; this is new for  $s \geq 4$ . More precisely, we establish that if  $f : [N] \rightarrow [-1, 1]$  is a function with  $\|f\|_{U^{s+1}[N]} \geq \delta$  then there is a bounded-complexity  $s$ -step nilsequence  $F(g(n)\Gamma)$  which correlates with  $f$ , where the bounds on the complexity and correlation depend only on  $s$  and  $\delta$ . From previous results, this conjecture implies the Hardy-Littlewood prime tuples conjecture for any linear system of finite complexity.

## CONTENTS

1. Introduction	1
2. Strategy of the proof	4
3. Basic notation	8
4. The polynomial formulation of $\text{GI}(s)$	11
5. Taking ultralimits	14
6. Nilcharacters and symbols in one and several variables	18
7. A more detailed outline of the argument	31
8. A variant of Gowers's Cauchy-Schwarz argument	35
9. Frequencies and representations	39
10. Linear independence and the sunflower lemma	48
11. Obtaining bracket-linear behaviour	58
12. Building a nilobject	64
13. The symmetry argument	71
Appendix A. Basic theory of ultralimits	77
Appendix B. Polynomial algebra	83
Appendix C. Lifting linear nilsequences to polynomial ones	90
Appendix D. Equidistribution theory	92
Appendix E. Some basic properties of nilcharacters and symbols	99
Appendix F. A linearisation result from additive combinatorics	113
References	114

## 1. INTRODUCTION

The purpose of this paper is to establish the general case of a conjecture named the *Inverse Conjecture for the Gowers norms* by the first two authors in [23, Conjecture 8.3]. If  $N$  is a (typically large) positive integer then we write  $[N] := \{1, \dots, N\}$ . For each integer  $s \geq 1$  the inverse conjecture  $\text{GI}(s)$ , whose statement we recall shortly, describes the structure of 1-bounded functions  $f : [N] \rightarrow \mathbb{C}$  whose  $(s+1)^{\text{st}}$

Gowers norm  $\|f\|_{U^{s+1}[N]}$  is large. These conjectures together with a good deal of motivation and background to them are discussed in [19, 21, 23]. The conjectures GI(1) and GI(2) have been known for some time, the former being a straightforward application of Fourier analysis, and the latter being the main result of [21] (see also [51] for the characteristic 2 analogue). The case GI(3) was also recently established by the authors in [28]. The aim of the present paper is to establish the remaining cases GI( $s$ ) for  $s \geq 3$ , in particular reestablishing the results in [28].

We begin by recalling the definition of the Gowers norms. If  $G$  is a finite abelian group,  $d \geq 1$  is an integer, and  $f : G \rightarrow \mathbb{C}$  is a function then we define

$$\|f\|_{U^d(G)} := (\mathbb{E}_{x, h_1, \dots, h_d \in G} \Delta_{h_1} \dots \Delta_{h_d} f(x))^{1/2^d}, \quad (1.1)$$

where  $\Delta_h f$  is the multiplicative derivative

$$\Delta_h f(x) := f(x+h) \overline{f(x)}$$

and  $\mathbb{E}_{x \in X} f(x) := \frac{1}{|X|} \sum_{x \in X} f(x)$  denotes the average of a function  $f : X \rightarrow \mathbb{C}$  on a finite set  $X$ . Thus for instance we have

$$\|f\|_{U^2(G)} := \left( \mathbb{E}_{x, h_1, h_2 \in G} f(x) \overline{f(x+h_1)} \overline{f(x+h_2)} f(x+h_1+h_2) \right)^{1/4}.$$

One can show that  $U^d(G)$  is indeed a norm on the functions  $f : G \rightarrow \mathbb{C}$  for any  $d \geq 2$ , though we will not need this fact here.

In this paper we will be concerned with functions on  $[N]$ , which is not quite a group. To define the Gowers norms of a function  $f : [N] \rightarrow \mathbb{C}$ , set  $G := \mathbb{Z}/\tilde{N}\mathbb{Z}$  for some integer  $\tilde{N} \geq 2^d N$ , define a function  $\tilde{f} : G \rightarrow \mathbb{C}$  by  $\tilde{f}(x) = f(x)$  for  $x = 1, \dots, N$  and  $\tilde{f}(x) = 0$  otherwise, and set

$$\|f\|_{U^d[N]} := \|\tilde{f}\|_{U^d(G)} / \|1_{[N]}\|_{U^d(G)},$$

where  $1_{[N]}$  is the indicator function of  $[N]$ . It is easy to see that this definition is independent of the choice of  $\tilde{N}$ . One could take  $\tilde{N} := 2^d N$  for definiteness if desired.

The *Inverse conjecture for the Gowers  $U^{s+1}[N]$ -norm*, abbreviated as GI( $s$ ), posits an answer to the following question.

**Question 1.1.** *Suppose that  $f : [N] \rightarrow \mathbb{C}$  is a function bounded in magnitude by 1, and let  $\delta > 0$  be a positive real number. What can be said if  $\|f\|_{U^{s+1}[N]} \geq \delta$ ?*

Note that in the extreme case  $\delta = 1$  one can easily show that  $f$  is a phase polynomial, namely  $f(n) = e(P(n))$  for some polynomial  $P$  of degree at most  $s$ . Furthermore, if  $f$  correlates with a phase polynomial, that is to say if  $|\mathbb{E}_{n \in [N]} f(n) \overline{e(P(n))}| \geq \delta$ , then it is easy to show that  $\|f\|_{U^{s+1}[N]} \geq c(\delta)$ . It is natural to ask whether the converse is also true - does a large Gowers norm imply correlation with a polynomial phase function? Surprisingly, the answer is no, as was observed by Gowers [16] and, in the related context of *multiple recurrence*, somewhat earlier by Furstenberg and Weiss [13, 14]. The work of Furstenberg-Weiss and Conze-Lesigne [10] draws attention to the role of homogeneous spaces  $G/\Gamma$  of nilpotent Lie groups, and subsequent work of Host and Kra [36] provides a link, in an ergodic-theoretic context, between these spaces and certain seminorms with a formal similarity to the Gowers norms under discussion here. Later work of Bergelson, Host and Kra [4] highlights the role of a class of functions arising from these spaces  $G/\Gamma$  called *nilsequences*. The inverse conjecture for the Gowers norms, first formulated precisely in [23, §8], postulates

that this class of functions (which contains the polynomial phases) represents the full set of obstructions to having large Gowers norm.

We now recall that precise formulation. Recall that an  $s$ -step nilmanifold is a manifold of the form  $G/\Gamma$ , where  $G$  is a connected, simply-connected nilpotent Lie group of step at most  $s$  (i.e. all  $s+1$ -fold commutators of  $G$  are trivial), and  $\Gamma$  is a discrete, cocompact<sup>1</sup> subgroup of  $G$ .

**Conjecture 1.2** (GI( $s$ )). *Let  $s \geq 0$  be an integer, and let  $0 < \delta \leq 1$ . Then there exists a finite collection  $\mathcal{M}_{s,\delta}$  of  $s$ -step nilmanifolds  $G/\Gamma$ , each equipped with some smooth Riemannian metric  $d_{G/\Gamma}$  as well as constants  $C(s,\delta), c(s,\delta) > 0$  with the following property. Whenever  $N \geq 1$  and  $f : [N] \rightarrow \mathbb{C}$  is a function bounded in magnitude by 1 such that  $\|f\|_{U^{s+1}[N]} \geq \delta$ , there exists a nilmanifold  $G/\Gamma \in \mathcal{M}_{s,\delta}$ , some  $g \in G$  and a function  $F : G/\Gamma \rightarrow \mathbb{C}$  bounded in magnitude by 1 and with Lipschitz constant at most  $C(s,\delta)$  with respect to the metric  $d_{G/\Gamma}$  such that*

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F(g^n x)}| \geq c(s,\delta).$$

We remark that there are many equivalent ways to reformulate this conjecture. For instance, instead of working with a finite family  $\mathcal{M}_{s,\delta}$  of nilmanifolds, one could work with a single nilmanifold  $G/\Gamma = G_{s,\delta}/\Gamma_{s,\delta}$ , by taking the Cartesian product of all the nilmanifolds in the family. Other reformulations include an equivalent formulation using polynomial nilsequences rather than linear ones (see Conjecture 4.5) and an ultralimit formulation (see Conjecture 5.3). One can also formulate the conjecture using bracket polynomials, or local polynomials; see [21] for a discussion of these equivalences in the  $s = 2$  case.

Let us briefly review the known partial results on this conjecture:

- (i) GI(0) is trivial.
- (ii) GI(1) follows from a short Fourier-analytic computation.
- (iii) GI(2) was established about five years ago in [21], building on work of Gowers [16].
- (iv) GI(3) was established, quite recently, in [28].
- (v) In the extreme case  $\delta = 1$  one can easily show that  $f(n) = e(P(n))$  for some polynomial  $P$  of degree at most  $s$ , and every such function is an  $s$ -step nilsequence by a direct construction. See, for example, [21] for the case  $s = 2$ .
- (vi) In the almost extremal case  $\delta \geq 1 - \varepsilon_s$ , for some  $\varepsilon_s > 0$ , one may see that  $f$  correlates with a phase  $e(P(n))$  by adapting arguments first used in the theoretical computer-science literature [1].
- (vii) The analogue of GI( $s$ ) in ergodic theory (which, roughly speaking, corresponds to the asymptotic limit  $N \rightarrow \infty$  of the theory here; see [37] for further discussion) was formulated and established in [36], work done independently of the work of Gowers (see also the earlier paper [35]). This work was the first place in the literature to link objects of Gowers-norm type (associated to functions on a measure-preserving system  $(X, T, \mu)$ ) with flows on nilmanifolds, and the subsequent paper [4] was the first work to underline the importance of *nilsequences*. The formulation of GI( $s$ ) by the first two authors in [23] was very strongly influenced by these works. For the closely related problem of analysing multiple ergodic averages, the

---

<sup>1</sup>A subgroup  $\Gamma$  of a topological group  $G$  is *cocompact* if the quotient space  $G/\Gamma$  is compact.

relevance of flows on nilmanifolds was earlier pointed out in [13, 14, 47], building upon earlier work in [10]. See also [34, 60] for related work on multiple averages and nilmanifolds in ergodic theory.

- (viii) The analogue of  $\text{GI}(s)$  in finite fields of large characteristic was established by ergodic-theoretic methods in [5, 56].
- (ix) A weaker “local” version of the inverse theorem (in which correlation takes place on a subprogression of  $[N]$  of size  $\sim N^{c_s}$ ) was established by Gowers [17]. This paper provided a good deal of inspiration for our work here.
- (x) The converse statement to  $\text{GI}(s)$ , namely that correlation with a function of the form  $n \mapsto F(g^n x)$  implies that  $f$  has large  $U^{s+1}[N]$ -norm, is also known. This was first established in [21, Proposition 12.6], following arguments of Host and Kra [36] rather closely. A rather simple proof of this result is given in [28, Appendix G].

The main result of this paper is a proof of Conjecture 1.2:

**Theorem 1.3.** *For any  $s \geq 3$ , the inverse conjecture for the  $U^{s+1}[N]$ -norm,  $\text{GI}(s)$ , is true.*

By combining this result with the previous results in [23, 26] we obtain a quantitative Hardy-Littlewood prime tuples conjecture for all linear systems of finite complexity; in particular, we now have the expected asymptotic for the number of primes  $p_1 < \dots < p_k \leq X$  in arithmetic progression, for every fixed positive integer  $k$ . We refer to [23] for further discussion, as we have nothing new to add here regarding these applications. Several further applications of the  $\text{GI}(s)$  conjectures are given in [11, 27].

## 2. STRATEGY OF THE PROOF

The proof of Theorem 1.3 is long and complicated, but broadly speaking it follows the strategy laid out in previous works [16, 17, 21, 28, 51]. We induct on  $s$ , assuming that  $\text{GI}(s-1)$  has already been established and using this to prove  $\text{GI}(s)$ . To explain the argument, let us first summarise the main steps taken in [28] in order to deduce  $\text{GI}(3)$ , the inverse theorem for the  $U^4$ -norm, from  $\text{GI}(2)$ , the inverse theorem for the  $U^3$  norm (established in [21]). Once this is done we will explain some of the extra difficulties involved in handling the general case. For a more extensive (but informal) discussion of the proof strategy, see [29]. Once we set up some technical machinery, we will also be able to give a more detailed description of the strategy in §7.

Here, then, is an overview of the argument in [28].

- (i) (Apply induction) If  $\|f\|_{U^4[N]} \gg 1$  then, for many  $h$ ,  $\|\Delta_h f\|_{U^3[N]} \gg 1$  and so  $\Delta_h f$  correlates with a 2-step nilsequence  $\chi_h$ .
- (ii) (Nilcharacter decomposition)  $\chi_h$  may be decomposed as a sum of a special type of nilsequence called a *nilcharacter*, essentially by a Fourier decomposition. For the sake of illustration, these 2-step nilcharacters may be supposed to have the form

$$\chi_h(n) = e(\{\alpha_h n\} \beta_h n),$$

although these are not quite nilcharacters due to the discontinuous nature of the fractional part function  $x \mapsto \{x\}$ , and in any event a general 2-step nilcharacter will be modeled by a linear combination of such “bracket

quadratic monomials”, rather than by a single such monomial (see [21] for further discussion).

- (iii) (Rough linearity) The fact that  $\Delta_h f$  correlates with  $\chi_h$  forces  $\chi_h$  to behave weakly linearly in  $h$ . To get a feel for why this is so, suppose that  $|f| \equiv 1$ ; then we have the cocycle identity

$$\Delta_{h+k} f(n) = \Delta_h f(n+k) \Delta_k f(n).$$

To capture something like the same behaviour in the much weaker setting where  $\Delta_h f$  correlates with  $\chi_h$ , we use an extraordinary argument of Gowers [16] relying on the Cauchy-Schwarz inequality. Roughly speaking, the information obtained is of the form

$$\chi_{h_1} \chi_{h_2} \sim \chi_{h_3} \chi_{h_4} \quad \text{modulo lower order terms} \quad (2.1)$$

for many  $h_1, h_2, h_3, h_4$  with  $h_1 + h_2 = h_3 + h_4$ .

- (iv) (Furstenberg-Weiss) An argument of Furstenberg and Weiss [14] is adapted in order to study (2.1). The quantitative distribution theory of nilsequences developed in [24] is a major input here. It is concluded that we may assume that the frequency  $\beta_h$  does not actually depend on  $h$ . Note that this step appeared for the first time in the proof of GI(3); it did not feature in the proof of GI(2) in [21].
- (v) (Linearisation) A similar argument allows one to then assert that

$$\alpha_{h_1} + \alpha_{h_2} \approx \alpha_{h_3} + \alpha_{h_4} \pmod{1} \quad (2.2)$$

for many  $h_1, h_2, h_3, h_4$  with  $h_1 + h_2 = h_3 + h_4$ .

- (vi) (Additive Combinatorics) By arguments from additive combinatorics related to the Balog-Szemerédi-Gowers theorem [2, 16] and Freiman’s theorem, as well as some geometry of numbers, we may then assume that  $\alpha_h$  varies “bracket-linearly” in  $h$ , thus

$$\alpha_h = \gamma_1 \{\eta_1 h\} + \cdots + \gamma_d \{\eta_d h\}. \quad (2.3)$$

Up to top order, then, the nilcharacter  $\chi_h(n)$  can now be assumed to take the form  $e(\psi(h, n, n))$ , where  $\psi$  is “bracket-multilinear”; it is a sum of terms such as  $\{\gamma \{\eta h\} n\} \beta n$ .

- (vii) (Symmetry argument) The bracket multilinear form  $\psi$  obeys an additional symmetry property. This is a reflection of the identity  $\Delta_h \Delta_k f = \Delta_k \Delta_h f$ , but transferring this to the much weaker setting in which we merely have correlation of  $\Delta_h f$  with  $\chi_h$  requires another appeal to Gowers’ Cauchy-Schwarz argument from (iii). In fact, the key point is to look at the second order terms in (2.1).
- (viii) (Integration) Assuming this symmetry, one is able to express

$$\chi_h(n) \sim \Theta(n+h) \overline{\Theta'(n)}$$

for some bracket cubic functions  $\Theta, \Theta'$ , which morally take the form

$$\Theta(n), \Theta'(n) \sim e(\psi(n, n, n)/3)$$

(for much the same reason that  $x^3/3$  is an antiderivative of  $x^2$ ). Thus we morally have

$$\Delta_h f(n) \sim \Theta(n+h) \overline{\Theta'(n)}$$

- (ix) (Construction of a nilsequence) Any bracket cubic form like  $e(\psi(n, n, n))$  “comes from” a 3-step nilmanifold; this construction is accomplished in [28] in a rather *ad hoc* manner.
- (x) From here, one can analyse lower order terms by the induction hypothesis GI(2). This is a relatively easy matter.

Let us now discuss the argument of this paper in the light of each point of this outline. A more detailed outline is given in §7. Assume that GI( $s - 1$ ) has been established.

- (i) (Apply induction) If  $\|f\|_{U^{s+1}[N]} \gg 1$  then, for many  $h$ ,  $\|\Delta_h f\|_{U^s[N]} \gg 1$  and so  $\Delta_h f$  correlates with an  $(s - 1)$ -step nilsequence  $\chi_h$ . This is straightforward (see §7).
- (ii) (Nilcharacter decomposition)  $\chi_h$  may be decomposed into nilcharacters; this is fairly straightforward as well. It is somewhat reassuring to think of  $\chi_h(n)$  as having the form  $e(\psi_h(n))$ , where  $\psi_h(n)$  is a bracket polynomial “of degree  $s - 1$ ”, but we will not be working explicitly with bracket polynomials much in this paper, except as motivation and as a source of examples. One of the main challenges one is faced with during an attempt to prove GI(4) by a direct generalisation of our arguments from [28] is the fact that already bracket cubic polynomials are rather complicated to deal with and can take different forms such as  $\{\alpha n\}\{\beta n\}\gamma n$  and  $\{\{\alpha n\}\beta n\}\gamma n$ .

Instead of objects such as  $e(\alpha n\{\beta n\})$ , then, we will work with the rather more abstract notion of a *symbol*. This notion, which is fairly central to our paper, is defined and discussed in §6. One additional technical point is worth mentioning here. This is the fact that  $e(\alpha n\{\beta n\})$  (say) cannot be realised as a nilsequence  $F(g^n \Gamma)$  with  $F$  *continuous*, and therefore the distributional results of [24] do not directly apply. In [28] these discontinuities could be understood quite explicitly, but here we take a different approach: we decompose  $G/\Gamma$  into  $D$  pieces using a smooth partition of unity for some  $D = O(1)$ , and then work instead with the (smooth)  $\mathbb{C}^D$ -valued nilsequence consisting of these pieces.

We discuss this device more fully in §6, but we emphasise that this is a technical device and the reader is advised not to give this particular aspect of the proof too much attention.

- (iii) (Rough linearity)  $\chi_h$  varies roughly linearly in  $h$ ; this is another fairly straightforward modification of the arguments of Gowers, already employed in [28], which is performed in §8.
- (iv) (Furstenberg-Weiss) This proceeds along similar lines to the corresponding argument in [28] but is, in a sense, rather easier once one has developed the device of  $\mathbb{C}^D$ -valued nilsequences, which allow one to remain in the smooth category; this is accomplished in §11, after a substantial amount of preparatory material in §9, §10 and Appendix D.
- (v) (Linearisation) This is also quite similar to the corresponding argument in [28], and is performed in §11. In both of parts (iv) and (v), the “bracket calculus” from [28] is replaced by the more conceptual “symbol calculus” developed in Appendix E.
- (vi) (Additive Combinatorics) The additive combinatorial input is much the same as in [28]. For the convenience of the reader we sketch it in Appendix F.

- (vii) (Construction of a nilsequence) Our argument differs quite substantially from that in [28] at this point. The  $s$ -step nilobject, which is now a two-variable object  $\chi(h, n)$ , is constructed *before* the symmetry argument and in a more conceptual manner. This may be compared with the rather *ad hoc* approach taken in [21, 28], where various bracket polynomials were merely exhibited as arising from nilsequences. We perform this construction in §12.
- (viii) (Symmetry argument) We replace  $\chi(h, n)$  with an equivalent nilcharacter  $\tilde{\chi}(h, n, \dots, n)$  where  $\tilde{\chi}$  is a nilcharacter in  $s$  variables, that is symmetric in the last  $s - 1$  variables. The symmetry argument given in §13 shows that  $\tilde{\chi}(h, n, \dots, n)$  is equivalent to  $\tilde{\chi}(n, h, \dots, n)$ . Again the key idea in the analysis is to look at the second order terms in (2.1).
- (ix) (Integration) With the symmetry in hand, we can use the calculus of multilinear nilcharacters essentially express  $\tilde{\chi}(h, n, \dots, n)$  as the derivative of an expression which is roughly of the form  $\tilde{\chi}(n, \dots, n)/s$ ; see §13 for details.
- (x) The final step of the argument is relatively straightforward, as before; see §7.

In our previous paper [28] it was already rather painful to keep proper track of such notions as “many” and “correlates with”. Here matters are even worse, and so to organise the above tasks it turns out to be quite convenient to first take an ultralimit of all objects being studied, effectively placing one in the setting of *nonstandard analysis*. This allows one to easily import results from infinitary mathematics, notably the theory of Lie groups and basic linear algebra, into the finitary setting of functions on  $[N]$ . In §5 and Appendix A we review the basic machinery of ultralimits that we will need here; we will not be exploiting any particularly advanced aspects of this framework. The reader does not really need to understand the ultrafilter language in order to comprehend the basic structure of the paper, provided that he/she is happy to deal with concepts like “dense” and “correlates with” in a somewhat informal way, resembling the way in which analysts actually talk about ideas with one another (and, in fact, analogous to the way we wrote this paper). It is possible to go through the paper and properly quantify all of these notions using appropriate parameters  $\delta$  and (many) growth functions  $\mathcal{F}$ . This would have the advantage of making the paper on some level comprehensible to the reader with an absolute distrust of ultrafilters, and it would also remove the dependence on the axiom of choice and in principle provide explicit but very poor bounds. However it would cause the argument to be significantly longer, and the notation would be much bulkier.

Our exposition will be as follows. We will begin by spending some time introducing the ultrafilter language and then, motivated by examples, the notions of nilsequence, nilcharacter and symbol. Once that is done we will, in §7, give the high-level argument for Theorem 1.3; this consist of detailing points (i), (ii) and (x) of the outline above and giving proper statements of the other main points.

The discussion above concerning points (iii), (iv), (v) and (vi) has been simplified for the sake of exposition. In actual fact, these points are dealt with together by a kind of iterative loop, in which more and more bracket-linear structure is placed on the nilcharacters  $\chi_h(n)$  by cycling from (iii) to (vi) repeatedly.

We remark that a quite different approach using ultrafilters to the structural theory of the Gowers norms is in the process of being carried out in [52, 53, 54];

this seems related to the work of Host and Kra, whereas our work ultimately derives from the work of Gowers.

We also make the minor remark that our proof of  $\text{GI}(s)$  is restricted to the case  $s \geq 3$  for minor technical reasons. In particular, we take advantage of the non-trivial nature of the degree  $s - 2$  “lower order terms” in the Gowers Cauchy-Schwarz argument (Proposition 7.3) in the symmetry argument step; and we will also observe that the various “smooth” and “periodic” error terms arising from the equidistribution theory in Appendix D are of degree 1 and thus negligible compared with the main terms in the analysis, which are of degree  $s - 1$ . The arguments can be modified to give a proof of  $\text{GI}(2)$ , although this proof would basically be a notationally intensive repackaging of the arguments in [21].

*Acknowledgements.* BG was, for some of the period during which this work was carried out, a fellow of the Radcliffe Institute at Harvard. He is very grateful to the Radcliffe Institute for providing excellent working conditions. TT is supported by NSF Research Award DMS-0649473, the NSF Waterman award and a grant from the MacArthur Foundation. TZ is supported by ISF grant 557/08, an Alon fellowship and a Landau fellowship of the Taub foundation. All three authors are very grateful to the University of Verona for allowing them to use classrooms at Canazei during a week in July 2009. This work was largely completed during that week.

### 3. BASIC NOTATION

We write  $\mathbb{N} := \{0, 1, 2, \dots\}$  for the natural numbers, and  $\mathbb{N}^+ := \{1, 2, \dots\}$  for the positive natural numbers. Given two integers  $N, M$ , we write  $[N, M]$  for the discrete interval  $[N, M] := \{n : N \leq n \leq M\}$ . We also make the abbreviations  $[N] := [1, N]$ , and  $[-N] := [-N, N]$ . If  $x$  is a real number, we write  $x \bmod 1$  for the associated residue class in the unit circle  $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ , and write  $x = y \bmod 1$  if  $x$  and  $y$  differ by an integer.

We will rely frequently on the following two elementary functions: the *fundamental character*  $e : \mathbb{R} \rightarrow \mathbb{C}$  (or  $e : \mathbb{T} \rightarrow \mathbb{C}$ ) defined by

$$e(x) := e^{2\pi i x},$$

and the *signed fractional part function*<sup>2</sup>  $\{x\} : \mathbb{R} \rightarrow I_0$ , where  $I_0$  is the *fundamental domain*

$$I_0 := \{x \in \mathbb{R} : -1/2 < x \leq 1/2\}$$

and  $\{x\}$  is the unique real number in  $I_0$  such that  $x = \{x\} \bmod 1$ . We will often rely on the identity

$$e(x) = e(\{x\}) = e(x \bmod 1)$$

without further comment.

For technical reasons, we will need to manipulate vector-valued complex quantities in a manner analogous to scalar complex quantities. If  $v = (v_i)_{i=1}^D$  and  $w = (w_i)_{i=1}^{D'}$  are vectors in  $\mathbb{C}^D$  and  $\mathbb{C}^{D'}$  respectively then we form the *tensor product*  $v \otimes w \in \mathbb{C}^{DD'}$  by the formula

$$v \otimes w := (v_1 w_1, \dots, v_D w_{D'})$$

---

<sup>2</sup>The signed fractional part will be slightly more convenient to work with than the unsigned fractional part, as it is equal to the identity near the origin.



and the *complex conjugate*  $\bar{v} \in \mathbb{C}^D$  by the formula

$$\bar{v} := (\bar{v}_1, \dots, \bar{v}_D).$$

Similarly, if  $X$  is some set and  $f : X \rightarrow \mathbb{C}^D$  and  $g : X \rightarrow \mathbb{C}^{D'}$  are functions then we write  $f \otimes g : X \rightarrow \mathbb{C}^{DD'}$  for the function defined by  $(f \otimes g)(x) := f(x) \otimes g(x)$ , and similarly define  $\bar{f} : X \rightarrow \mathbb{C}^D$ .

If  $G = (G, +)$  is an additive group,  $k \in \mathbb{N}$ ,  $\vec{g} = (g_1, \dots, g_k) \in G^k$ , and  $\vec{a} = (a_1, \dots, a_k) \in \mathbb{Z}^k$ , we define the dot product

$$\vec{a} \cdot \vec{g} := a_1 g_1 + \dots + a_k g_k.$$

Given a set  $H$  in an additive group, define an *additive quadruple* in  $H$  to be a quadruple  $(h_1, h_2, h_3, h_4) \in H$  with  $h_1 + h_2 = h_3 + h_4$ . The number of additive quadruples in  $H$  is known as the *additive energy* of  $H$  and is denoted  $E(H)$ .

A map  $\phi : H \rightarrow G$  from  $H$  to another additive group  $G$  is said to be a *Freiman homomorphism* if it preserves additive quadruples, i.e. if  $\phi(h_1) + \phi(h_2) = \phi(h_3) + \phi(h_4)$  for all additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ .

Given a multi-index  $\vec{d} = (d_1, \dots, d_k) \in \mathbb{N}^k$ , we write  $|\vec{d}| := d_1 + \dots + d_k$ .

We now briefly review and clarify some standard notation from group theory.

When we do not assume a group  $G$  to be abelian, we will always write  $G$  multiplicatively:  $G = (G, \cdot)$ . However, when dealing with abelian groups, we reserve the right to use additive notation instead.

We view an  $n$ -tuple  $(a_1, \dots, a_n)$  of labels as a finite ordered set with the ordering  $a_1 < \dots < a_n$ . If  $A = (a_1, \dots, a_n)$  is a finite ordered set and  $(g_a)_{a \in A}$  are a collection of group elements in a multiplicative group  $G$ , we define the ordered products

$$\prod_{a \in A} g_a := g_{a_1} \dots g_{a_n}, \quad \prod_{i=1}^n g_i := g_1 \dots g_n \quad \text{and} \quad \prod_{i=n}^1 g_i := g_n \dots g_1$$

for any  $n \geq 0$ , with the convention that the empty product is the identity. We extend this notation to infinite products under the assumption that all but finitely many of the factors are equal to the identity.

Given a subset  $A$  of a group  $G$ , we let  $\langle A \rangle$  denote the subgroup of  $G$  generated by  $A$ . Given a family  $(H_i)_{i \in I}$  of subgroups of  $G$ , we write  $\bigvee_{i \in I} H_i$  for the smallest subgroup of  $G$  that contains all of the  $H_i$ .

Given two elements  $g, h$  of a multiplicative group  $G$ , we define the *commutator*

$$[g, h] := g^{-1} h^{-1} g h.$$

We write  $H \leq G$  to denote the statement that  $H$  is a subgroup of  $G$ . If  $H, K \leq G$ , we let  $[H, K]$  be the subgroup generated by the commutators  $[h, k]$  with  $h \in H$  and  $k \in K$ , thus  $[H, K] = \langle \{[h, k] : h \in H, k \in K\} \rangle$ .

If  $r \geq 1$  is an integer and  $g_1, \dots, g_r \in G$ , we define an  $(r-1)$ -fold *iterated commutator* of  $g_1, \dots, g_r$  inductively by declaring  $g_1$  to be the only 0-fold iterated commutator of  $g_1$ , and for  $r > 1$  defining an  $(r-1)$ -fold iterated commutator to be any expression of the form  $[w, w']$ , where  $w$  and  $w'$  are  $(s-1)$ -fold and  $(s'-1)$ -fold commutators of  $g_{i_1}, \dots, g_{i_s}$  and  $g_{i'_1}, \dots, g_{i'_{s'}}$  respectively, where  $s, s' \geq 1$  are such that  $s + s' = r$ , and  $\{i_1, \dots, i_s\} \cup \{i'_1, \dots, i'_{s'}\} = \{1, \dots, r\}$  is a partition of  $\{1, \dots, r\}$  into two classes. Thus for instance  $[[g_3, g_1], [g_2, g_4]]$  and  $[g_2, [g_1, [g_3, g_4]]]$  are 3-fold iterated commutators of  $g_1, \dots, g_4$ .

The following lemma will be useful for computing commutator groups.

**Lemma 3.1.** *Let  $H = \langle A \rangle, K = \langle B \rangle$  be normal subgroups of a nilpotent group  $G$  that are generated by sets  $A \subset H, B \subset K$  respectively. Then  $[H, K]$  is normal, and is also the subgroup generated by the  $i + j - 1$ -fold iterated commutators of  $a_1, \dots, a_i, b_1, \dots, b_j$  with  $a_1, \dots, a_i \in A, b_1, \dots, b_j \in B$  and  $i, j \geq 1$ .*

*Proof.* The normality of  $[H, K]$  follows from the identity

$$g[H, K]g^{-1} = [gHg^{-1}, gKg^{-1}].$$

It is then clear that  $[H, K]$  contains the group generated by the iterated commutators of elements in  $A, B$  that involve at least one element from each. The converse follows inductively using the identities

$$[x, y] = [y, x]^{-1}, \quad [xy, z] = [x, z][x, z], y[y, z] \quad \text{and} \quad [x, y^{-1}] = [y, x][y, x], y^{-1}. \quad (3.1)$$

This concludes the proof.  $\square$

As a corollary of the above lemma, we have the distributive law

$$\left[ \bigvee_{i \in I} H_i, \bigvee_{j \in J} K_j \right] = \bigvee_{i \in I, j \in J} [H_i, K_j]$$

whenever  $(H_i)_{i \in I}, (K_j)_{j \in J}$  are families of normal subgroups of a nilpotent group  $G$ .

If  $H \triangleleft G$  is a normal subgroup of  $G$ , and  $g \in G$ , we use  $g \bmod H$  to denote the coset representative  $gH$  of  $g$  in  $G/H$ . For instance,  $g = g' \bmod H$  if  $gH = g'H$ .

At various stages in the paper we will need the (discrete) *Baker-Campbell-Hausdorff formula* in the following weak form:

$$g_1^{n_1} g_2^{n_2} = g_2^{n_2} g_1^{n_1} \prod_a g_a^{P_a(n_1, n_2)} \quad (3.2)$$

for all  $g_1, g_2$  in a nilpotent group  $G$  and all integers  $n_1, n_2$ , where  $g_a$  ranges over all iterated commutators of  $g_1, g_2$  that involve at least one copy of each (note from nilpotency that there are only finitely many non-trivial  $g_a$ ), with the  $a$  ordered in some arbitrary fashion, and  $P_a : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$  are polynomials. Furthermore, if  $g_a$  involves  $d_1$  copies of  $g_1$  and  $d_2$  copies of  $g_2$ , then  $P_a$  has degree at most  $d_1$  in the  $n_1$  variable and  $d_2$  in the  $n_2$  variable.

Let  $G$  be a connected, simply connected, nilpotent Lie group (or *nilpotent Lie group* for short). Then we denote the Lie algebra of  $G$  as  $\log G$ . As is well known (see e.g. [7]), the exponential map  $\exp : \log G \rightarrow G$  is a homeomorphism, inverted by the logarithm map  $\log : G \rightarrow \log G$ , and we can then define the exponentiation operation  $g^t$  for any  $g \in G$  and  $t \in \mathbb{R}$  by the formula

$$g^t := \exp(t \log g).$$

There is a continuous version of the Baker-Campbell-Hausdorff formula:

$$g_1^{t_1} g_2^{t_2} = g_2^{t_2} g_1^{t_1} \prod_a g_a^{P_a(t_1, t_2)} \quad (3.3)$$

for all  $t_1, t_2 \in \mathbb{R}$  and  $g_1, g_2 \in G$ , where  $P_a$  are the polynomials occurring in (3.2). We also observe the variant formulae

$$(g_1 g_2)^t = g_1^t g_2^t \prod_a g_a^{Q_a(t)}$$

for some polynomials  $Q_a$  and all  $t \in \mathbb{R}$ ,  $g_1, g_2 \in G$ , and

$$\exp(t_1 \log g_1 + t_2 \log g_2) = g_1^{t_1} g_2^{t_2} \prod_a g_a^{R_a(t_1, t_2)}$$

for some further polynomials  $R_a$  and all  $t_1, t_2 \in \mathbb{R}$ ,  $g_1, g_2 \in G$ . We refer to all of these formulæ collectively as *the Baker-Campbell-Hausdorff formula*.

If  $A$  is a subset of a nilpotent Lie group  $G$ , we let  $\langle A \rangle_{\mathbb{R}}$  be the smallest connected Lie subgroup of  $G$  containing  $A$ , or more explicitly

$$\langle A \rangle_{\mathbb{R}} := \langle \{a^t : a \in A; t \in \mathbb{R}\} \rangle.$$

Equivalently,  $\log \langle A \rangle_{\mathbb{R}}$  is the Lie algebra generated by  $\log A$ .

A *lattice* of a nilpotent Lie group  $G$  is a discrete cocompact subgroup  $\Gamma$  of  $G$ . Thus for instance, we see from (3.2) that for any finite set  $A$  in  $G$ ,  $\langle A \rangle$  will be a cocompact subgroup of  $\langle A \rangle_{\mathbb{R}}$ , and will thus be a lattice if  $\langle A \rangle$  is discrete.

A connected Lie subgroup  $H$  of  $G$  is said to be *rational* with respect to  $\Gamma$  if  $\Gamma \cap H$  is cocompact in  $H$ . For instance, if  $G = \mathbb{R}^2$ ,  $\Gamma$  is the standard lattice  $\mathbb{Z}^2$ , and  $\alpha \in \mathbb{R}$ , then the connected Lie subgroup  $H := \{(x, \alpha x) : x \in \mathbb{R}\}$  is rational if and only if  $\alpha$  is rational.

**FURTHER NOTATION.** Here is a list of further notation used in the paper for reference, together with the place in the paper where each piece is defined and discussed.

$\text{poly}(H_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$	polynomial maps from one filtered group $H_{\mathbb{N}}$ to $G_{\mathbb{N}}$	6.18
$\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$	polynomial maps with the degree filtration	6.18
$\text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow G_{\mathbb{N}^k})$	polynomial maps with the multidegree filtration	6.18
$\text{poly}(\mathbb{Z}_{\text{DR}} \rightarrow G_{\text{DR}})$	polynomial maps with the degree-rank filtration	6.18
$L^{\infty}(\Omega \rightarrow \overline{\mathbb{C}}^D)$	bounded limit functions to ${}^*\mathbb{C}^d$	(A.1)
$L^{\infty}(\Omega \rightarrow \overline{\mathbb{C}}^w)$	bounded limit functions (also $L^{\infty}(\Omega)$ )	(A.1)
$\text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$	bd'd limit functions with bounded Lipschitz constant	5.1
$\text{Nil}^d([N])$	nilsequences of degree $\leq d$ on $[N]$	5.2
$\text{Nil}^{\subset J}(\Omega)$	nilsequences of degree $\subset J$	6.19
$\Xi^d([N])$	space of degree $d$ nilcharacters on $[N]$	6.1
$\Xi_{\text{Multi}}^{(d_1, \dots, d_k)}(\Omega)$	multidegree nilcharacters	6.19
$\Xi_{\text{DR}}^{(d, r)}(\Omega)$	degree-rank nilcharacters	6.19
$\text{Symb}^d([N])$	equiv. classes of degree $d$ nilcharacters in $\Xi^d([N])$	6.6
$\text{Symb}_{\text{Multi}}^{(d_1, \dots, d_k)}(\Omega)$	equiv. classes of multidegree nilcharacters	6.22
$\text{Symb}_{\text{DR}}^{(d, r)}(\Omega)$	equiv. classes of degree-rank nilcharacters	6.22
$G^{\vec{D}}, G^{\vec{D}, \leq (s-1, r_*)}$	universal nilpotent Lie group of degree-rank $(s-1, r_*)$	9.1
$\text{Horiz}_i(G)$	$i$ 'th horizontal space of $G$	9.6
$\text{Taylor}_i(g)$	$i$ 'th horizontal Taylor coefficient of a polynomial map	9.6
$(\vec{D}, \eta, \mathcal{F})$	total frequency representation of a nilcharacter	9.11

#### 4. THE POLYNOMIAL FORMULATION OF $\text{GI}(s)$

The inverse conjecture  $\text{GI}(s)$ , Conjecture 1.2, has been formulated using *linear* nilsequences  $F(g^n x \Gamma)$ . This is largely for compatibility with the earlier paper [23] of the first two authors on linear equations in primes, where this form of the conjecture was stated in precisely this form as Conjecture 8.3. Subsequently, however, it was

discovered that it is more natural to deal with a somewhat more general class of object called a *polynomial nilsequence*  $F(g(n)\Gamma)$ . This is particularly so when it comes to discussing the distributional properties of nilsequences, as was done in [24]. Thus, we shall now recast the inverse conjecture in terms of polynomial nilsequences, which is the formulation we will work with throughout the rest of the paper.

Let us first recall the definition of a polynomial nilsequence of degree  $d$ .

**Definition 4.1** (Polynomial nilsequence). Let  $G$  be a (connected, simply-connected) nilpotent Lie group. By a *filtration*  $G_{\mathbb{N}} = (G_i)_{i \in \mathbb{N}}$  of degree  $\leq d$  we mean a nested sequence  $G \supseteq G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_{d+1} = \{\text{id}\}$  with the property that  $[G_i, G_j] \subseteq G_{i+j}$  for all  $i, j \geq 0$ , adopting the convention that  $G_i = \{\text{id}\}$  for all  $i > d$ . By a *polynomial sequence* adapted to  $G_{\mathbb{N}}$  we mean a map  $g : \mathbb{Z} \rightarrow G$  such that  $\partial_{h_i} \dots \partial_{h_1} g \in G_i$  for all  $i \geq 0$  and  $h_1, \dots, h_i \in \mathbb{Z}$ , where  $\partial_h g(n) := g(n+h)g(n)^{-1}$ . Write  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  for the collection of all such polynomial sequences.

Let  $\Gamma \leq G$  be a lattice in  $G$  (i.e. a discrete and cocompact subgroup), so that the quotient  $G/\Gamma$  is a nilmanifold, and assume that each of the  $G_i$  are *rational* subgroups (i.e.  $\Gamma_i := \Gamma \cap G_i$  is a cocompact subgroup of  $G_i$ ). We refer to the pair  $G/\Gamma = (G/\Gamma, G_{\mathbb{N}})$  as a *filtered nilmanifold*. A *polynomial orbit*  $\mathcal{O} : \mathbb{Z} \rightarrow G/\Gamma$  is a sequence of the form  $\mathcal{O}(n) := g(n)\Gamma$ , where  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ ; we let  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  denote the space of all such polynomial orbits. If  $F : G/\Gamma \rightarrow \mathbb{C}$  is a 1-bounded, Lipschitz function then the sequence  $F \circ \mathcal{O} = (F(g(n)\Gamma))_{n \in \mathbb{Z}}$  is called a *polynomial nilsequence* of degree  $d$ .

The subscripts  $\mathbb{N}$  will become more relevant later in this paper, when we start filtering nilpotent groups and nilmanifolds by other index sets  $I$  than the natural numbers  $\mathbb{N}$ . Note that we do not require  $G_0$  or  $G_1$  to equal  $G$ ; this freedom will be convenient for some minor technical reasons, although ultimately it will not enlarge the space of polynomial nilsequences.

Let us give the basic examples of nilsequences and polynomials:

**Example 4.2** (Linear nilsequences are polynomial nilsequences). Let  $G$  be a  $d$ -step nilpotent Lie group, and let  $\Gamma$  be a lattice of  $G$ . Then, as is well known (see e.g. [7]), the *lower central series filtration* defined by  $G_0 = G_1 := G$ ,  $G_2 := [G, G_1]$ ,  $G_3 := [G, G_2], \dots, G_{d+1} := [G, G_d] = \{\text{id}\}$  is a filtration on  $G$ . Using the Baker-Campbell-Hausdorff formula (3.3) it is not difficult to show that the lower central series filtration is rational with respect to  $\Gamma$ , so the nilmanifold  $G/\Gamma$  becomes a filtered nilmanifold. If  $g(n) := g_1^n g_0$  for some  $g_0, g_1 \in G$ , then  $\partial_{h_1} g(n) = g_1^{h_1}$  and  $\partial_{h_i} \dots \partial_{h_1} g(n) = \text{id}$  for  $i \geq 2$ : therefore  $g$  is a polynomial sequence, and so every linear orbit  $n \mapsto g^n x$  with  $g \in G$  and  $x \in G/\Gamma$  is a polynomial orbit also. As a consequence we see that every  $d$ -step linear nilsequence  $n \mapsto F(g^n x)$  is automatically a polynomial nilsequence of degree  $\leq d$ .

**Example 4.3** (Polynomial phases are polynomial nilsequences). Let  $d \geq 0$  be an integer. Then we can give the unit circle  $\mathbb{T}$  the structure of a degree  $\leq d$  filtered nilmanifold by setting  $G := \mathbb{R}$  and  $\Gamma := \mathbb{Z}$ , with  $G_i := \mathbb{R}$  for  $i \leq d$  and  $G_i := \{0\}$  for  $i > d$ . This is clearly a filtered nilmanifold. If  $\alpha_0, \dots, \alpha_d$  are real numbers, then the polynomial  $P(n) := \alpha_0 + \dots + \alpha_d n^d$  is then polynomial with respect to this filtration, with  $n \mapsto P(n) \bmod 1$  being a polynomial orbit in  $\mathbb{T}$ . Thus, for any Lipschitz function  $F : \mathbb{T} \rightarrow \mathbb{C}$ , the sequence  $n \mapsto F(P(n))$  is a polynomial

nilsequence of degree  $\leq d$ ; in particular, the polynomial phase  $n \mapsto e(P(n))$  is a polynomial nilsequence.

**Example 4.4** (Combinations of monomials are polynomials). By Corollary B.4, we see that if  $G = (G, (G_i)_{i \in \mathbb{N}})$  is a filtered group of degree  $\leq d$ , then any sequence of the form

$$n \mapsto \prod_{j=1}^k g_j^{P_j(n)},$$

in which  $g_j \in G_{d_j}$  for some  $d_j \in \mathbb{N}$ , and  $P_j : \mathbb{Z} \rightarrow \mathbb{R}$  is a polynomial of degree  $\leq d_j$ , will be a polynomial map. Thus for instance

$$n \mapsto g_d^{\binom{n}{d}} \dots g_2^{\binom{n}{2}} g_1^n g_0$$

is a polynomial map whenever  $g_j \in G_j$  for  $j = 0, \dots, d$ . In fact, all polynomial maps can be expressed in such a fashion via a *Taylor expansion*; see Lemma B.9.

We will give several further examples and properties of polynomial maps and polynomial nilsequences in §6.

As a consequence of Example 4.2, the following variant of the inverse conjecture  $\text{GI}(s)$  is ostensibly weaker than that stated in the introduction.

**Conjecture 4.5** ( $\text{GI}(s)$ , polynomial formulation). *Let  $s \geq 0$  be an integer, and let  $0 < \delta \leq 1$ . Then there exists a finite collection  $\mathcal{M}_{s,\delta}$  of filtered nilmanifolds  $G/\Gamma = (G/\Gamma, G_{\mathbb{N}})$ , each equipped with some smooth Riemannian metric  $d_{G/\Gamma}$  as well as constants  $C(s, \delta), c(s, \delta) > 0$  with the following property. Whenever  $N \geq 1$  and  $f : [N] \rightarrow \mathbb{C}$  is a function bounded in magnitude by 1 such that  $\|f\|_{U^{s+1}[N]} \geq \delta$ , there exists a filtered nilmanifold  $G/\Gamma \in \mathcal{M}_{s,\delta}$ , some  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  and a function  $F : G/\Gamma \rightarrow \mathbb{C}$  bounded in magnitude by 1 and with Lipschitz constant at most  $C(s, \delta)$  with respect to the metric  $d_{G/\Gamma}$  such that*

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F(g(n)\Gamma)}| \geq c(s, \delta).$$

It turns out that this conjecture is actually *equivalent* to Conjecture 1.2; we shall prove this equivalence in Appendix C. We remark that, though it might seem odd to put a non-trivial part of the proof of our main theorem in an appendix, we would rather encourage the reader to regard the proof of Conjecture 4.5 as our main theorem. The rationale behind this is that everything that is done with linear nilsequences  $F(g^n x \Gamma)$  in [23] could have been done equally well, and perhaps more naturally, with polynomial nilsequences  $F(g(n)\Gamma)$ . Further remarks along these lines were made in the introduction to our earlier paper [28], where the polynomial formulation was emphasised from the outset. Here, however, we have felt a sense of duty to formally complete the programme outlined in [23].

Henceforth we shall refer simply to a *nilsequence*, rather than a polynomial nilsequence.

In §6 we will need to generalise the notion of a (polynomial) nilsequence by allowing more exotic filtrations  $G_I$  on the group  $G$ , indexed by more complicated index sets  $I$  than the natural numbers  $\mathbb{N}$ . In particular, we shall introduce the *multidegree filtration*, which allows us to define nilsequences of several variables, as well as the *degree-rank* filtration which provides a finer classification of polynomial sequences than merely the degree. We will discuss these using examples, and then develop a more unified theory that contains all three.

## 5. TAKING ULTRALIMITS

The inverse conjecture, Conjecture 4.5, is a purely finitary statement, involving functions on a finite set  $[N] = \{1, \dots, N\}$  of integers. As such, it is natural to look for proofs of this conjecture which are also purely finitary, and much of the previous literature on these types of problems is indeed of this nature.

However there is a very notable exception, namely the portion of the literature that exploits the *Furstenberg correspondence principle* between combinatorial problems and ergodic theory. See [12] for the original application to Szemerédi’s theorem, or [56] for a more recent application to Gowers norms over finite fields. Here we use a somewhat different type of limit object, namely an *ultralimit*. We are certainly not the first to employ ultralimits (a.k.a. *nonstandard analysis*) in additive number theory; see for example [40].

The ultralimit formalism allows us to convert a “finitary” or “standard” statement such as Conjecture 4.5 into an equivalent statement concerning *limit objects*, constructed as ultralimits of standard objects. This procedure is closely related to the use of the *transfer principle* in nonstandard analysis, but we have elected to eschew the language of nonstandard analysis in order to reduce confusion, instead focusing on the machinery of ultralimits.

Here is a brief and somewhat vague list of the advantages of using the ultralimit approach.

- Pigeonholing arguments are straightforward (due to the fact that a limit function taking finitely many values is constant);
- Book-keeping of constants: one can talk rigorously about such concepts as “bounded” functions without a need to quantify the bounds;
- One may make rigorous sense of such statements as “the function  $f : [N] \rightarrow \mathbb{C}$  and the function  $g : [N] \rightarrow \mathbb{C}$  are equivalent modulo degree  $s$  nilsequences”.
- In the infinitary context one may easily perform *rank reduction* arguments in which one seeks to find the “minimal bounded-complexity” representation of a given system.

There are also some drawbacks of the approach:

- It becomes quite difficult to extract any quantitative bounds from our results, in particular we do not give explicit bounds on the constant  $c(s, \delta)$  or on the complexity of the nilsequence in Conjecture 1.2 or Conjecture 4.5. It is in principle possible to expand the ultralimit proof into a standard proof, but the bounds are quite poor (of Ackermann type) due to the repeated use of “rank reduction arguments” and other highly iterative schemes that arise in the conversion of ultralimit arguments to standard ones. For further discussion of the relation of ultralimit analysis to finitary analysis see [55, §1.3, §1.5].
- The language of ultrafilters adds one more layer of notational complexity to an already notationally-intensive paper; however, there are gains to be made elsewhere, most notably in eliminating many quantitative constants (e.g.  $\varepsilon$ ,  $N$ ) and growth functions (e.g.  $\mathcal{F}$ ).

**LIMIT FORMULATION OF  $\text{GI}(s)$ .** The basic notation and theory of ultralimits are reviewed in Appendix A. We now use this formalism to convert the inverse conjecture,  $\text{GI}(s)$ , into an equivalent statement formulated in the framework of

ultralimits. We first consider a limit version of the concept of a Lipschitz function on a nilmanifold. For technical reasons we will need to consider vector-valued functions, taking values in  $\mathbb{C}^D$  or  $\overline{\mathbb{C}}^D$  rather than  $\mathbb{C}$  or  $\overline{\mathbb{C}}$ .

**Definition 5.1** (Lipschitz functions). Let  $G/\Gamma$  be a standard nilmanifold, and let  $D \in \mathbb{N}^+$  be standard.

- We let  $\text{Lip}(G/\Gamma \rightarrow \mathbb{C}^D)$  be the space of standard Lipschitz functions  $F : G/\Gamma \rightarrow \mathbb{C}^D$ . (Here we endow the compact manifold  $G/\Gamma$  with a smooth metric in an arbitrary fashion; the exact choice of metric is not relevant.)
- We let  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$  be the space of bounded limit functions  $F : *G/\Gamma \rightarrow \overline{\mathbb{C}}^D$  whose Lipschitz constant is bounded (or equivalently,  $F$  is an ultralimit of uniformly bounded functions  $F_n : G/\Gamma \rightarrow \mathbb{C}^D$  with uniformly bounded Lipschitz constant).
- We let  $\text{Lip}(*G/\Gamma \rightarrow \overline{S^{2D-1}})$  be the functions in  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$  that take values in the (limit) complex sphere

$$\overline{S^{2D-1}} := \{z \in \overline{\mathbb{C}}^D : |z| = 1\}.$$

- We write

$$\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^\omega) := \bigcup_{D \in \mathbb{N}^+} \text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$$

and

$$\text{Lip}(*G/\Gamma \rightarrow \overline{S}^\omega) := \bigcup_{D \in \mathbb{N}^+} \text{Lip}(*G/\Gamma \rightarrow \overline{S^{2D-1}}).$$

We will often abbreviate these spaces as  $\text{Lip}(G/\Gamma)$  or  $\text{Lip}(*G/\Gamma)$  when the range of the functions involved is not relevant to the discussion.

*Remark.* As  $G/\Gamma$  is compact, we see from the Arzelà-Ascoli theorem that  $\text{Lip}(G/\Gamma \rightarrow \mathbb{C}^D)$  is locally compact in the  $L^\infty(G/\Gamma \rightarrow \mathbb{C}^D)$  topology. As a consequence, if we embed  $\text{Lip}(G/\Gamma \rightarrow \mathbb{C}^D)$  into  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$  in the obvious manner, then the former is a dense subspace of the latter in the (standard) uniform topology, in the sense that for every  $F \in \text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$  and every standard  $\varepsilon > 0$  there exists  $F' \in \text{Lip}(G/\Gamma \rightarrow \mathbb{C}^D)$  such that  $|F(x) - F'(x)| \leq \varepsilon$  for all  $x \in *G/\Gamma$ .

*Remark.* Observe that the spaces  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^D)$  and  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^\omega)$  are vector spaces over  $\overline{\mathbb{C}}$ . The spaces  $\text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}}^\omega)$  and  $\text{Lip}(*G/\Gamma \rightarrow \overline{S}^\omega)$  are also closed under tensor product (as defined in §3). All the spaces defined in Definition 5.1 are closed under complex conjugation.

Using the above notion, we can define the limit version of a (polynomial) nilsequence.

**Definition 5.2** (Nilsequence). Let  $s \geq 0$  be standard. A *nilsequence* of degree  $\leq s$  is any limit function  $\psi : *\mathbb{Z} \rightarrow *\mathbb{C}$  of the form  $\psi(n) := F(g(n)\Gamma)$ , where  $G/\Gamma = (G/\Gamma, G_\mathbb{N})$  is a standard filtered nilmanifold of degree  $\leq s$ ,  $g : *\mathbb{Z} \rightarrow *G$  is a limit polynomial sequence (i.e. an ultralimit of polynomial sequences  $g_n : \mathbb{Z} \rightarrow G$ ), and  $F \in \text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}})$ .

Given any limit subset  $\Omega$  of  $*\mathbb{Z}$ , we denote the space of degree  $d$  nilsequences, restricted to  $\Omega$ , as  $\text{Nil}^d(\Omega) = \text{Nil}^d(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ ; this is a subset of  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ . We

write  $\text{Nil}^d(\Omega \rightarrow \overline{\mathbb{C}}^D)$  for the nilsequences that take values in  $\overline{\mathbb{C}}^D$ ; this is a subspace (over  $\overline{\mathbb{C}}$ ) of  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D)$ . We make the technical remark that  $\text{Nil}^d(\Omega)$  is a  $\sigma$ -limit set, since one can express this space as the union, over all standard  $M$  and dimensions  $D$ , of the nilsequences taking values in  $\overline{\mathbb{C}}^D$  arising from a nilmanifold of “complexity”  $M$  and a Lipschitz function of constant at most  $M$ , where one defines the complexity of a nilmanifold in some suitable fashion. In particular, the limit selection lemma in Corollary A.12 can be applied to this set.

We also define the Gowers uniformity norm  $\|f\|_{U^{s+1}[N]}$  of an ultralimit  $f = \lim_{\mathbf{n} \rightarrow p} f_{\mathbf{n}}$  of standard functions  $f_{\mathbf{n}} : [N_{\mathbf{n}}] \rightarrow \mathbb{C}$  in the usual limit fashion

$$\|f\|_{U^{s+1}[N]} := \lim_{\mathbf{n} \rightarrow p} \|f_{\mathbf{n}}\|_{U^{s+1}[N_{\mathbf{n}}]}.$$

If  $f$  is vector-valued instead of scalar valued, say  $f = (f_1, \dots, f_d)$ , then we define the uniformity norm by the formula

$$\|f\|_{U^{s+1}[N]} := \left( \sum_{i=1}^d \|f_i\|_{U^{s+1}[N]}^{2^{s+1}} \right)^{1/2^{s+1}}.$$

(The exponent  $2^{s+1}$  is not important here, but has some very slight aesthetic advantages over other equivalent formulations of the vector-valued norm.)

The ultralimit formulation of  $\text{GI}(s)$  can then be given as follows:

**Conjecture 5.3** (Ultralimit formulation of  $\text{GI}(s)$ ). *Let  $s \geq 0$  be standard and  $N \geq 1$  be a limit natural number. Suppose that  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}})$  is such that  $\|f\|_{U^{s+1}[N]} \gg 1$ . Then  $f$  correlates with a degree  $\leq s$  nilsequence on  $[N]$ .*

See Definition A.7 for the definition of *correlation* in this context.

We now show why, for any fixed standard  $s$ , Conjecture 5.3 is equivalent to its more traditional counterpart, Conjecture 4.5.

*Proof of Conjecture 5.3 assuming Conjecture 4.5.* Let  $f$  be as in Conjecture 5.3. We may normalise the bounded function  $f$  to be bounded by 1 in magnitude throughout. By hypothesis, there exists a standard  $\delta > 0$  such that  $\|f\|_{U^{s+1}[N]} \geq \delta$ . Writing  $N$  and  $f$  as the ultralimits of  $N_{\mathbf{n}}$ ,  $f_{\mathbf{n}}$  respectively for some  $f_{\mathbf{n}} : [N_{\mathbf{n}}] \rightarrow \mathbb{C}$  bounded in magnitude by 1, and applying Conjecture 4.5, we conclude that for  $\mathbf{n}$  sufficiently close to  $p$ , we have the correlation bound

$$|\mathbb{E}_{n_{\mathbf{n}} \in [N_{\mathbf{n}}]} f_{\mathbf{n}}(n_{\mathbf{n}}) \overline{F_{\mathbf{n}}(g_{\mathbf{n}}(n_{\mathbf{n}})\Gamma_{\mathbf{n}})}| \geq c(s, \delta) > 0$$

where  $G_{\mathbf{n}}/\Gamma_{\mathbf{n}}, g_{\mathbf{n}}, x_{\mathbf{n}}, F_{\mathbf{n}}$  are as in Conjecture 1.2. Writing  $G/\Gamma, g, x, F$  for the ultralimits of  $G_{\mathbf{n}}/\Gamma_{\mathbf{n}}, g_{\mathbf{n}}, x_{\mathbf{n}}, F_{\mathbf{n}}$  respectively, we thus have

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F(g(n)^*\Gamma)}| \gg 1.$$

By the pigeonhole principle (cf. Appendix A), we see that  $G/\Gamma$  is a standard degree  $\leq s$  nilmanifold, while  $g : \mathbb{Z} \rightarrow *G$  and  $x \in G/\Gamma$  remain limit objects. The limit function  $F$  lies in  $\text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}})$  by construction, and the claim follows.  $\square$

*Proof of Conjecture 4.5 assuming Conjecture 5.3.* Observe (from the theory of Mal'cev bases [48]) that there are only countably many degree  $\leq s$  nilmanifolds  $G/\Gamma$  up to isomorphism, which we may enumerate as  $G_{\mathbf{n}}/\Gamma_{\mathbf{n}}$ . We endow each of these nilmanifolds arbitrarily with some smooth Riemannian metric  $d_{G_{\mathbf{n}}/\Gamma_{\mathbf{n}}}$ .

Suppose for contradiction that Conjecture 4.5 failed. Carefully negating all the quantifiers, we may thus find a  $\delta > 0$ , a sequence  $N_{\mathbf{n}}$  of standard integers, and a



function  $f_{\mathbf{n}} : [N_{\mathbf{n}}] \rightarrow \mathbb{C}$  bounded in magnitude by 1 with  $\|f_{\mathbf{n}}\|_{U^{s+1}[N]} \geq \delta$ , such that

$$|\mathbb{E}_{n_{\mathbf{n}} \in [N_{\mathbf{n}}]} f_{\mathbf{n}}(n_{\mathbf{n}}) \overline{F(g(n_{\mathbf{n}})\Gamma_{\mathbf{n}'})})| \leq 1/\mathbf{n} \quad (5.1)$$

whenever  $\mathbf{n}' \leq \mathbf{n}$ ,  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G_{\mathbf{n}'}_{\mathbb{N}})_{\mathbb{N}})$ , and  $F : G_{\mathbf{n}'}_{\mathbb{N}}/\Gamma_{\mathbf{n}'} \rightarrow \mathbb{C}$  is bounded in magnitude by 1 and has a Lipschitz constant of at most  $\mathbf{n}$  with respect to  $d_{G_{\mathbf{n}'}_{\mathbb{N}}/\Gamma_{\mathbf{n}'}}$ .

On the other hand, viewing  $f$  as a bounded limit function, we can apply Conjecture 5.3 and conclude that there exists a standard filtered nilmanifold  $G/\Gamma$  with some smooth Riemannian metric  $d_{G/\Gamma}$ , a limit polynomial  $g : \mathbb{Z} \rightarrow *G$ , and some ultralimit  $F \in \text{Lip}(*G/\Gamma \rightarrow \overline{\mathbb{C}})$  of functions  $F_{\mathbf{n}} : G/\Gamma \rightarrow \mathbb{C}$  with uniformly bounded Lipschitz norm, such that

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F(g(n)*\Gamma)}| \geq \varepsilon$$

for some standard  $\varepsilon > 0$ .

By construction,  $G/\Gamma$  is isomorphic to  $G_{\mathbf{n}_0}/\Gamma_{\mathbf{n}_0}$  for some  $\mathbf{n}_0$ , so we may assume without loss of generality that  $G/\Gamma = G_{\mathbf{n}_0}/\Gamma_{\mathbf{n}_0}$ ; since all smooth Riemannian metrics on a compact manifold are equivalent, we can also assume that  $d_{G/\Gamma} = d_{G_{\mathbf{n}_0}/\Gamma_{\mathbf{n}_0}}$ . We may also normalise  $F$  to be bounded in magnitude by 1. But this contradicts (5.1) for  $\mathbf{n}$  sufficiently large, and the claim follows.

Thus, to establish Theorem 1.3, it will suffice to establish Conjecture 5.3 for  $s \geq 3$ . This is the objective of the remainder of the paper.

*Remark.* We transformed the finitary linear inverse conjecture, Conjecture 1.2, into a nonstandard polynomial formulation, Conjecture 5.3, via the finitary polynomial inverse conjecture, Conjecture 4.5. One can also swap the order of these equivalences, transforming the finitary linear inverse conjecture into a nonstandard linear formulation by arguing as above, and then transforming the latter into a nonstandard polynomial formulation by using Proposition C.2. Of course the two arguments are essentially equivalent.

Conjecture 5.3 is trivial when  $N$  is bounded, since every function in  $L^\infty[N]$  is then a nilsequence of degree at most  $s$ . For the remainder of the paper we shall thus adopt the convention that  $N$  denotes a fixed *unbounded* limit integer.

To conclude this section we reformulate Conjecture 4.5 by introducing the important notion of *bias*.

**Definition 5.4** (Bias and correlation). Let  $\Omega$  be a limit finite subset of  $\mathbb{Z}$ , and let  $d \in \mathbb{N}$ . We say that  $f, g \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$  *d-correlate* if we have

$$|\mathbb{E}_{n \in \Omega} f(n) \otimes \overline{g(n)} \otimes \psi(n)| \gg 1$$

for some degree  $d$  nilsequence  $\psi \in \text{Nil}^d(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ . We say that  $f$  is *d-biased* if  $f$  *d-correlates* with the constant function 1, and *d-unbiased* otherwise.

With this definition, Conjecture 5.3 can be reformulated in the following manner.

**Conjecture 5.5** (Limit formulation of GI(s), II). *Let  $s \geq 0$  be standard. Suppose that  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}})$  is such that  $\|f\|_{U^{s+1}[N]} \gg 1$ . Then  $f$  is  $s$ -biased.*

From previous literature, we see that Conjecture 5.5 has already been proven for  $s \leq 2$ ; we need to establish it for all  $s \geq 3$ . We also make the basic remark that while the conjecture is only phrased for scalar-valued functions  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}})$ , it automatically generalises to vector-valued functions  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}}^\omega)$ , since if a vector-valued function  $f$  has large  $U^{s+1}[N]$  norm, then so does one of its components.

Finally we remark that the converse implication is known.

**Proposition 5.6** (Converse GI( $s$ ), ultralimit formulation). *Let  $s \geq 0$  be standard. Suppose that  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}})$  is  $\leq s$ -biased. Then  $\|f\|_{U^{s+1}[N]} \gg 1$ .*

*Proof.* This follows from [21, Proposition 12.6], [23, §11], or [28, Proposition 1.4], transferred to the ultralimit setting in the usual fashion.  $\square$

## 6. NILCHARACTERS AND SYMBOLS IN ONE AND SEVERAL VARIABLES

Conjecture 5.3 asserts that a function in  $L^\infty([N] \rightarrow \overline{\mathbb{C}})$  on an unbounded interval  $[N]$  correlates with a degree  $\leq s$  nilsequence. For inductive reasons, it is useful to observe that this conclusion implies a strengthened version of itself, in which  $f$  correlates with a special type of degree  $\leq s$  nilsequence, namely a degree  $s$  *nilcharacter*. A nilcharacter is a special type of nilsequence and should be thought of, very roughly speaking, as a generalisation of characters  $e(\alpha n)$  in the degree 1 setting, or objects such as  $e(\alpha n \{\beta n\})$  in the degree 2 setting; these were crucial in our paper on GI(3) [28], although the notation there was slightly different in some minor ways. See [29] for further informal discussion of nilcharacters.

In the  $s = 1$  case, a nilcharacter is essentially (ignoring constants) the same thing as a linear phase function  $n \mapsto e(\xi n)$ , and the frequency  $\xi$  can be viewed as living in the Pontryagin dual of  ${}^*\mathbb{Z}$  (or, in some sense, of  $[N]$ , even though the latter set is not quite a locally compact abelian group). It will turn out that more generally, a degree  $s$  nilcharacter will have a “symbol” (analogous to the frequency  $\xi$ ) that takes values in a “higher order Pontryagin dual”  $\text{Symb}^s([N])$  of  $[N]$ ; this symbol can be interpreted as the “top order term” of a nilcharacter, for instance the symbol of the degree 3 nilcharacter  $n \mapsto e(\alpha n^3 + \beta n^2 + \gamma n + \delta)$  is basically<sup>3</sup>  $\alpha$ . This higher order dual obeys a number of pleasant algebraic properties, and the primary purpose of this section is to develop those properties.

There are various additional complications to be taken into account:

- We will require multidimensional generalisations of these concepts (think of the two-dimensional sequence  $(n_1, n_2) \mapsto e(\alpha n_1 \{\beta n_2\})$ ) together with appropriate notions of *multidegree* in order to make sense of “top-order” and “lower-order terms”;
- We will be dealing with  $\mathbb{C}^D$ -valued (or, rather,  $S^{2D-1}$ -valued) nilsequences rather than merely scalar ones. This is so that we may continue to work in the smooth category, as discussed in the introduction;
- The language of ultrafilters will be used.

Our main focus here will be on the first of these points. The second is largely a technicality, whilst the third is actually helpful in that the notion of symbol (for example) is rather clean and does not require discussion of complexity bounds.

**MOTIVATION AND ONE-DIMENSIONAL DEFINITIONS.** We now give the definitions of a (one-dimensional) nilcharacter and its symbol, and give a few examples. However, we will hold off for now on actually proving too much about these concepts, because we will shortly need to generalise these notions to a more abstract setting in which one also allows multidimensional nilcharacters, and nilcharacters that are attuned not just to a specific degree, but also to a specific “rank” inside that degree.

<sup>3</sup>This is an oversimplification; it would be more accurate to say that the symbol is given by  $\alpha$  modulo  ${}^*\mathbb{Z} + \mathbb{Q} + O(N^{-3})$ .

**Definition 6.1** (Nilcharacter). Let  $d \geq 0$  be a standard integer. A *nilcharacter*  $\chi$  of degree  $d$  on  $[N]$  is a nilsequence  $\chi(n) = F(\mathcal{O}(n)) = F(g(n)^*\Gamma)$  on  $[N]$  of degree  $\leq d$ , where the function  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}}^\omega)$  obeys two additional properties:

- $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{S}}^\omega)$  (thus  $|F| = 1$  pointwise, and hence  $|\chi| = 1$  pointwise also); and
- $F(g_dx) = e(\eta(g_d))F(x)$  for all  $x \in G/\Gamma$  and  $g_d \in G_d$ , where  $\eta : G_d \rightarrow \mathbb{R}$  is a continuous standard homomorphism which maps  $\Gamma_d$  to the integers (or equivalently,  $\eta$  is an element of the Pontryagin dual of the torus  $G_d/\Gamma_d$ ). We call  $\eta$  the *vertical frequency* of  $F$ .

The space of all nilcharacters of degree  $d$  on  $[N]$  is denoted  $\Xi^d([N])$ .

**Example 6.2.** When  $d = 1$ , the only examples of nilcharacters are the linear phases  $n \mapsto e(\alpha n + \beta)$  for  $\alpha, \beta \in {}^*\mathbb{R}$ .

**Example 6.3.** For any  $\alpha_0, \dots, \alpha_d \in {}^*\mathbb{R}$ , the function  $n \mapsto e(\alpha_0 + \dots + \alpha_d n^d)$  is a nilcharacter of degree  $\leq d$ . To see this, we set  $G/\Gamma$  to be the unit circle  $\mathbb{T}$  with the filtration  $G_i := \mathbb{R}$  for  $i \leq d$  and  $G_i := \{0\}$  for  $i > d$  (thus  $G/\Gamma$  is of degree  $d$ ), let  $g(n) := \alpha_0 + \dots + \alpha_d n^d$ , and let  $F(x) := e(x)$ . The vertical frequency  $\eta : \mathbb{R} \rightarrow \mathbb{R}$  is then just the identity function.

Now we give an instructive *near*-example of a nilcharacter. Let  $G$  be the free 2-step nilpotent Lie group on two generators  $e_1, e_2$ , thus

$$G := \langle e_1, e_2 \rangle_{\mathbb{R}} = \{e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}} : t_1, t_2, t_{12} \in \mathbb{R}\} \quad (6.1)$$

with the element  $[e_1, e_2]$  being central, but with no other relations between  $e_1, e_2$  and  $[e_1, e_2]$ . This is a degree  $\leq 2$  nilpotent group if we set  $G_0, G_1 := G$  and

$$G_2 := \langle [e_1, e_2] \rangle_{\mathbb{R}} = \{[e_1, e_2]^{t_{12}} : t_{12} \in \mathbb{R}\}.$$

We let

$$\Gamma := \langle e_1, e_2 \rangle = \{e_1^{n_1} e_2^{n_2} [e_1, e_2]^{n_{12}} : n_1, n_2, n_{12} \in \mathbb{Z}\}$$

be the discrete subgroup of  $G$  generated by  $e_1, e_2$ , then  $G/\Gamma$  is a degree  $\leq 2$  filtered nilmanifold, known as the *Heisenberg nilmanifold*, and elements of  $G/\Gamma$  can be uniquely expressed using the fundamental domain

$$G/\Gamma = \{e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}} \Gamma : t_1, t_2, t_{12} \in I_0 := (-1/2, 1/2]\}.$$

If we then set  $g : {}^*\mathbb{Z} \rightarrow {}^*G$  to be the limit polynomial sequence  $g(n) := e_2^{\beta n} e_1^{\alpha n}$  for some fixed  $\alpha, \beta \in {}^*\mathbb{R}$ , and let  $F : G/\Gamma \rightarrow \mathbb{C}$  be the function defined on the fundamental domain by the formula

$$F(e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}} \Gamma) := e(-t_{12}) \quad (6.2)$$

for  $t_1, t_2, t_{12} \in I_0$ , then one easily computes that

$$F(g(n)^*\Gamma) = e(\{\alpha n\}\beta n)$$

where  $\{\cdot\} : \mathbb{R} \rightarrow I_0$  is the signed fractional part function. The function  $n \mapsto e(\{\alpha n\}\beta n)$  is then *almost* a nilcharacter of degree 2, with vertical frequency given by the function  $\eta : [e_1, e_2]^{t_{12}} \mapsto -t_{12}$ . All the properties required to give a nilcharacter in Definition 6.1 are satisfied, save for one: the function  $F$  is not Lipschitz on all of  $G/\Gamma$ , but is instead merely *piecewise* Lipschitz, being discontinuous at some portions of the boundary of the fundamental domain. To put it another way, one can view  $n \mapsto e(\{\alpha n\}\beta n)$  as a *piecewise* nilcharacter of degree 2.

Indeed, a topological obstruction prevents one from constructing *any* scalar function  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{S^1})$  of unit magnitude on the Heisenberg nilmanifold with the above vertical frequency. By taking standard parts, we may assume that  $F$  comes from a standard Lipschitz function  $F : G/\Gamma \rightarrow S^1$  with the same vertical frequency. For any standard  $t \in [-1/2, 1/2]$ , consider the loop  $\gamma_t := \{e_1^t e_2^s \Gamma : s \in I_0\}$ . The image  $F(\gamma_t)$  of this loop lives on the unit circle and thus has a well-defined winding number (or degree). As this degree must vary continuously in  $t$  while remaining an integer, it is constant in  $t$ ; in particular,  $F(\gamma_{-1/2})$  and  $F(\gamma_{1/2})$  must have the same winding number. On the other hand, from the Baker-Campbell-Hausdorff formula (3.2) we see that

$$F(e_1^{1/2} e_2^s \Gamma) = F(e_1^{-1/2} e_2^s e_1 [e_1, e_2]^s \Gamma) = e(s) F(e_1^{-1/2} e_2^s \Gamma)$$

and so the winding number of  $F(\gamma_{1/2})$  is one larger than the winding number of  $F(\gamma_{-1/2})$ , a contradiction.

If however we allow ourselves to work with higher dimensions  $D$ , then this topological obstruction disappears. Indeed, let us take a smooth partition of unity  $1 = \sum_{k=1}^D \varphi_k^2(t, s)$  on  $\mathbb{T}^2$ , where  $D \in \mathbb{N}^+$  and each  $\varphi_k$  is supported in  $B_k \bmod \mathbb{Z}^2$ , where  $B_k$  is a ball of radius  $1/100$  (say) in  $\mathbb{R}^2$ . Then if we define  $F := (F_1, F_2, \dots, F_D)$ , where

$$F_k(e_1^t e_2^s [e_1, e_2]^u \Gamma) := \varphi_k(t, s) e(u) \quad (6.3)$$

whenever  $(t, s) \in {}^*B_k$  and  $u \in {}^*\mathbb{R}$ , with  $F_k = 0$  if no such representation of the above form exists, then one easily verifies that  $F$  lies in  $\text{Lip}^*(G/\Gamma \rightarrow \overline{S^{2D-1}})$  with the vertical frequency  $\eta$ , and so the vector-valued sequence  $\chi : n \mapsto F(g(n)^* \Gamma)$  is a nilcharacter of degree 2. A computation shows that each component  $\chi_k$  of this nilcharacter  $\chi = (\chi_1, \dots, \chi_D)$  takes the form

$$\chi_k(n) = e(\{\alpha n - \theta_k\} \beta n) \psi_k(n)$$

for some offset  $\theta_k \in {}^*\mathbb{R}$  and some degree 1 nilsequence  $\psi_k$ . Thus we see that  $\chi$  is in some sense “equivalent modulo lower order terms” with the bracket polynomial phase  $n \mapsto e(\{\alpha n\} \beta n)$ . We refer to the vector-valued nilsequence  $\chi$  as a *vector-valued smoothing* of the piecewise nilsequence  $n \mapsto e(\{\alpha n\} \beta n)$ ; we will informally refer to this smoothing operation several times in the sequel when discussing further examples of nilsequences that are associated with bracket polynomials.

Similar computations can be made in higher degree. For instance, bracket cubic phases such as  $n \mapsto e(\{\{\alpha n\} \beta n\} \gamma n)$  or  $n \mapsto e(\{\alpha n^2\} \beta n)$  with  $\alpha, \beta, \gamma \in {}^*\mathbb{R}$  can be viewed as near-examples of degree 3 nilcharacters (with the problem again being that  $F$  is discontinuous on the boundary of the fundamental domain), but there exist vector-valued smoothings of these phases which are genuine degree 3 nilcharacters. We will not detail these computations here, but they can essentially be found in [28, Appendix E]. More generally, one can view bracket polynomial phases of degree  $d$  as near-examples of nilcharacters of degree  $d$  that can be converted to genuine examples using vector-valued smoothings; this fact can be made precise using the machinery from [46], but we will not need this machinery here.

*Remark.* The above topological obstruction is quite annoying; it is the sole reason that we are forced to work with vector-valued functions. There are two other approaches to avoid this topological obstruction that we know of. One is to work with *piecewise* Lipschitz functions rather than Lipschitz functions. This

allows one in particular to build (piecewise) nilcharacters out of *bracket polynomials*. This is the approach taken in [28]; however, it requires one to develop a certain amount of “bracket calculus” to manipulate these polynomials, and some additional arguments are also needed to deal with the discontinuities at the edges of the piecewise components of the nilmanifold. Another approach is to work with randomly selected fundamental domains of the nilmanifold (cf. [20]) which eliminates topological obstructions, with the randomness being used to “average out” the effects of the boundary of the domain. While all three methods will eventually work for the purposes of establishing the inverse conjecture, we believe that the vector-valued approach introduces the least amount of artificial technicality.

By definition, every nilcharacter of degree  $d$  is a nilsequence of degree  $\leq d$ . The converse is far from being true; however, one can approximate nilsequences of degree  $\leq d$  as bounded linear combinations of nilcharacters of degree  $d$ . More precisely, we have the following lemma.

**Lemma 6.4.** *Let  $\psi \in \text{Nil}^d([N] \rightarrow \overline{\mathbb{C}})$  be a scalar nilsequence of degree  $d$ , and let  $\varepsilon > 0$  be standard. Then one can approximate  $\psi$  uniformly to error  $\varepsilon$  by a bounded linear combination (over  $\overline{\mathbb{C}}$ ) of the components of nilcharacters in  $\Xi^d([N])$ .*

*Proof.* Unpacking the definitions, it suffices to show that for every degree  $d$  filtered nilmanifold  $G/\Gamma$ , every  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}})$ , and every standard  $\varepsilon > 0$ , one can approximate  $F$  uniformly to error  $\varepsilon$  by a bounded linear combination of functions in the class  $\mathcal{F}(G/\Gamma)$  of components of standard Lipschitz functions  $F' \in \text{Lip}(G/\Gamma \rightarrow S^\omega)$  that have a vertical frequency in the sense of Definition 6.1.

By taking standard parts, we may assume that  $F$  is a standard Lipschitz function. Observe that  $\mathcal{F}(G/\Gamma)$  is closed under multiplication and complex conjugation. By the Stone-Weierstrass theorem, it thus suffices to show that  $\mathcal{F}(G/\Gamma)$  separates any two distinct points  $x, y \in G/\Gamma$ . If  $x, y$  do not lie in the same orbit of the  $G_d$ , then this is clear from a partition of unity (taking  $\eta = 0$ ). If instead  $x = g_d y$  for some  $g_d \in G_d$ , then the distinctness of  $x, y$  forces  $g_d \notin \Gamma_d$ , and hence by Pontryagin duality there exists a vertical frequency  $\eta$  with  $\eta(g_d) \neq 0$ . If one then builds a nilcharacter with this frequency (by adapting the vector-valued smoothing construction (6.3)) we obtain the claim.  $\square$

We remark that this lemma can also be proven, with better quantitative bounds, by Fourier-analytic methods: see [24, Lemma 3.7]. As a corollary of the lemma, we have the following.

**Corollary 6.5.** *Suppose that  $f \in L^\infty([N] \rightarrow \overline{\mathbb{C}}^\omega)$ . Then  $f$  is  $d$ -biased if and only if  $f$  correlates with a nilcharacter  $\chi \in \Xi^d([N])$ .*

It is easy to see that if  $\chi, \chi'$  are two nilcharacters of degree  $d$ , then the tensor product  $\chi \otimes \chi'$  and complex conjugate  $\overline{\chi}$  are also nilcharacters. If all nilcharacters were scalar, this would mean that the space  $\Xi^d([N])$  of degree  $d$  nilcharacters form a multiplicative abelian group. Unfortunately, nilcharacters can be vector-valued, and so this statement is not quite true. However, it becomes true if one only focuses on the “top order” behaviour of a nilcharacter. To isolate this behaviour, we adopt the following key definition.

**Definition 6.6** (Symbol). Let  $d \geq 0$ . Two nilcharacters  $\chi, \chi' \in \Xi^d([N])$  of degree  $d$  are *equivalent* if  $\chi \otimes \overline{\chi'}$  is equal on  $[N]$  to a nilsequence of degree  $\leq d-1$ . This can be shown to be an equivalence relation (see Lemma E.7); the equivalence class of a

nilcharacter  $\chi$  will be called the *symbol* of  $\chi$  and is denoted  $[\chi]_{\text{Symb}^d([N])}$ . The space of all such symbols will be denoted  $\text{Symb}^d([N])$ ; we will show later (see Lemma E.8) that this is an abelian multiplicative group.

When  $d = 1$ , two nilcharacters  $n \mapsto e(\alpha n + \beta)$  and  $n \mapsto e(\alpha' n + \beta')$  are equivalent if and only if  $\alpha - \alpha'$  is a limit integer, and  $\text{Symb}^1([N])$  is just  ${}^*\mathbb{T}$  in this case. However, the situation is more complicated in higher degree. To get some feel for this, consider two polynomial phases

$$\chi : n \mapsto e(\alpha_0 + \dots + \alpha_d n^d)$$

and

$$\chi' : n \mapsto e(\alpha'_0 + \dots + \alpha'_d n^d)$$

with  $\alpha_0, \dots, \alpha_d, \alpha'_0, \alpha'_d \in {}^*\mathbb{R}$ , and consider the problem of determining when  $\chi$  and  $\chi'$  are equivalent nilcharacters of degree  $d$ . Certainly this is the case if  $\alpha_d$  and  $\alpha'_d$  are equal, or differ by a limit integer. When  $d \geq 2$ , there are two further important cases in which equivalence occurs. The first is when  $\alpha'_d = \alpha_d + O(N^{-d})$ , because in this case the top degree component  $e((\alpha_d - \alpha'_d)n^d)$  of  $\chi\overline{\chi'}$  can be viewed as a Lipschitz function of  $n/2N \bmod 1$  (say) on  $[N]$  and is thus a 1-step nilsequence. The second is when  $\alpha'_d = \alpha_d + a/q$  for some standard rational  $q$ , since in this case the top degree component  $e((\alpha_d - \alpha'_d)n^d)$  of  $\chi\overline{\chi'}$  is periodic with period  $q$  and can thus be viewed as a Lipschitz function of  $n/q \bmod 1$  and is therefore again a 1-step nilsequence. We can combine all these cases together, and observe that  $\chi$  and  $\chi'$  are equivalent when  $\alpha'_d = \alpha_d + a/q + O(N^{-d}) \bmod 1$  for some standard rational  $a/q$ . It is possible to use the quantitative equidistribution theory of nilmanifolds (see [24]) to show that these are in fact the *only* cases in which  $\chi$  and  $\chi'$  are equivalent; this is a variant of the classical theorem of Weyl that a polynomial sequence is (totally) equidistributed modulo 1 if and only if at least one non-constant coefficients is irrational. In view of this, we see that  $\text{Symb}^d([N])$  contains  ${}^*\mathbb{R}/({}^*\mathbb{Z} + \mathbb{Q} + N^{-d}\overline{\mathbb{R}})$  as a subgroup, and the symbol of  $n \mapsto e(\alpha_0 + \dots + \alpha_d n^d)$  can be identified with

$$\alpha_d \bmod 1, \mathbb{Q}, O(N^{-d}) := \alpha + {}^*\mathbb{Z} + \mathbb{Q} + N^{-d}\overline{\mathbb{R}}.$$

However, the presence of bracket polynomials (suitably modified to avoid the topological obstruction mentioned earlier) means that when  $d \geq 2$ , that  $\text{Symb}^d([N])$  is somewhat larger than the above mentioned subgroup. We illustrate this with the following (non-rigorous) discussion. Take  $d = 2$  and consider two degree 2 nilcharacters  $\chi, \chi'$  of the form

$$\chi(n) \approx e(\{\alpha n\}\beta n + \gamma n^2)$$

and

$$\chi'(n) \approx e(\{\alpha' n\}\beta' n + \gamma' n^2)$$

for some  $\alpha, \beta, \gamma, \alpha', \beta', \gamma' \in {}^*\mathbb{R}$ , where we interpret the symbol  $\approx$  loosely to mean that  $\chi, \chi'$  are suitable vector-valued smoothings of the indicated bracket phases, of the type discussed earlier in this section. These may also involve some lower order nilsequences of degree 1.

As before, we consider the question of determining those values of  $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$  for which  $\chi$  and  $\chi'$  are equivalent. There are a number of fairly obvious ways in which equivalence can occur. For instance, by modifying the previous arguments, one can show that equivalence holds when  $\alpha = \alpha', \beta = \beta'$ , and  $\gamma - \gamma'$  is equal to a limit integer, a standard rational, or is equal to  $O(N^{-2})$ . Similarly, equivalence

occurs when  $\beta = \beta'$ ,  $\gamma = \gamma'$ , and  $\alpha - \alpha'$  is equal to a limit integer, a standard rational, or is equal to  $O(N^{-1})$ .

However, there are also some slightly less obvious ways in which equivalence can occur. Observe that the expression  $e(\{\alpha n\}\{\beta n\})$  is a Lipschitz function of the fractional parts of  $\alpha n$  and  $\beta n$  and is thus a (piecewise) nilsequence of degree 1 (and will become a genuine nilsequence after one performs an appropriate vector-valued smoothing). On the other hand, we have the obvious identity

$$e((\alpha n - \{\alpha n\})(\beta n - \{\beta n\})) = 1$$

since the exponent is the product of two (limit) integers. Expanding this out and rearranging, we obtain the (slightly imprecise) relation

$$e(\{\alpha n\}\beta n) \approx e(-\{\beta n\}\alpha n + \alpha\beta n^2) \quad (6.4)$$

where we again interpret  $\approx$  loosely to mean “after a suitable vector-valued smoothing, and ignoring lower order factors”. This gives an additional route for  $\chi$  and  $\chi'$  to be equivalent. A similar argument also gives the variant

$$e(\{\alpha n\}\beta n) \approx e\left(\frac{1}{2}\alpha\beta n^2\right)$$

whenever  $\alpha, \beta$  are *commensurate* in the sense that  $\alpha/\beta$  is a standard rational. We thus see that the notion of equivalence is in fact already somewhat complicated in degree 2, and the situation only becomes worse in higher degree. One can describe equivalence of bracket polynomials explicitly using *bracket calculus*, as developed in [46] (see also the earlier works [3, 30, 31, 32]), but this requires a fair amount of notation and machinery. Fortunately, in this paper we will be able to treat the notion of a symbol *abstractly*, without requiring an explicit description of the space  $\text{Symb}^d([N])$ .

**MORE GENERAL TYPES OF FILTRATION.** The notion of a one-dimensional polynomial  $n \mapsto \alpha_0 + \dots + \alpha_d n^d$  of degree  $\leq d$  can of course be generalised to higher dimensions. For instance, we have the notion of a multidimensional polynomial

$$(n_1, \dots, n_k) \mapsto \sum_{i_1, \dots, i_k \geq 0: i_1 + \dots + i_k \leq d} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k}$$

of degree  $\leq d$ . We also have the slightly different notion of a multidimensional polynomial

$$(n_1, \dots, n_k) \mapsto \sum_{i_1, \dots, i_k \geq 0: i_j \leq d_j \text{ for } 1 \leq j \leq k} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k}$$

of *multidegree*  $\leq (d_1, \dots, d_k)$  for some integers  $d_1, \dots, d_k \geq 0$ . We can unify these two concepts into the notion of a multi-dimensional polynomial

$$(n_1, \dots, n_k) \mapsto \sum_{(i_1, \dots, i_k) \in J} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k} \quad (6.5)$$

of *multidegree*  $\subset J$  for some finite *downset*  $J \subset \mathbb{N}^k$ , i.e. a finite set of tuples with the property that  $(i_1, \dots, i_k) \in J$  whenever  $(i_1, \dots, i_k) \in \mathbb{N}^k$  and  $i_j \leq i'_j$  for all  $j = 1, \dots, k$  for some  $(i'_1, \dots, i'_k) \in J$ . Thus for instance the two-dimensional polynomial

$$(h, n) \mapsto \alpha hn + \beta hn^2 + \gamma n^3$$

for  $\alpha, \beta, \gamma \in {}^*\mathbb{R}$  is of multidegree  $\subset J$  for

$$J := \{(0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 2)\},$$

and is also of multidegree  $\leq (1, 3)$  and of degree  $\leq 3$ . (One can view the downset  $J$  as a variant of the *Newton polytope* of the polynomial.)

In our subsequent arguments, we will need to similarly generalise the notion of a one-dimensional nilcharacter  $n \mapsto \chi(n)$  of degree  $\leq d$  to a multidimensional nilcharacter  $(n_1, \dots, n_k) \mapsto \chi(n_1, \dots, n_k)$  of degree  $\leq d$ , of multidegree  $\leq (d_1, \dots, d_k)$ , or of multidegree  $\subset J$ . We will define these concepts precisely in a short while, but we mention for now that the polynomial phase

$$(h, n) \mapsto e(\alpha hn + \beta hn^2 + \gamma n^3)$$

will be a two-dimensional nilcharacter of multidegree  $\subset J$ , multi-degree  $\leq (1, 3)$ , and degree  $\leq 3$  where  $J$  is as above. Moreover, variants of this phase, such as (a suitable vector-valued smoothing of)

$$(h, n) \mapsto e(\{\alpha_1 h\}\alpha_2 n + \{\{\beta_1 n\}\beta_2 h\}\beta_3 n + \{\gamma_1 n^2\}\gamma_2 n),$$

will also have the same multidegree and degree as the preceding example.

The multidegree of a nilcharacter  $\chi(n_1, \dots, n_k)$  is a more precise measurement of the complexity of  $\chi$  than the degree, because it separates the behaviour of the different variables  $n_1, \dots, n_k$ . We will also need a different refinement of the notion of degree, this time for a one-dimensional nilcharacter  $n \mapsto \chi(n)$ , which now separates the behaviour of different top degree components of  $\chi$ , according to their “rank”. Heuristically, the rank of such a component is the number of fractional part operations  $x \mapsto \{x\}$  that are needed to construct that component, plus one; thus for instance

$$n \mapsto e(\alpha n^3)$$

has degree 3 and rank 1,

$$n \mapsto e(\{\alpha n^2\}\beta n)$$

has degree 3 and rank 2 (after vector-valued smoothing),

$$n \mapsto e(\{\{\alpha n\}\beta n\}\gamma n)$$

has degree 3 and rank 3 (after vector-valued smoothing), and so forth. We will then need a notion of a nilcharacter  $\chi$  of *degree-rank*  $\leq (d, r)$ , which roughly speaking means that all the components used to build  $\chi$  either are of degree  $< d$ , or else are of degree exactly  $d$  but rank at most  $r$ . Thus for instance,

$$n \mapsto e(\{\alpha n\}\beta n + \gamma n^3)$$

has degree-rank  $\leq (3, 1)$  (after vector-valued smoothing), while

$$n \mapsto e(\{\alpha n\}\beta n + \gamma n^3 + \{\delta n^2\}\epsilon n)$$

has degree-rank  $\leq (3, 2)$  (after vector-valued smoothing), and

$$n \mapsto e(\{\alpha n\}\beta n + \gamma n^3 + \{\delta n^2\}\epsilon n + \{\{\mu n\}\nu n\}\rho n)$$

has degree-rank  $\leq (3, 3)$  (after vector-valued smoothing).

In order to make precise the notions of multidegree and degree-rank for nilcharacters, it is convenient to adopt an abstract formalism that unifies degree, multidegree, and degree-rank into a single theory. We need the following abstract definition.



**Definition 6.7** (Ordering). An *ordering*  $I = (I, \prec, +, 0)$  is a set  $I$  equipped with a partial ordering  $\prec$ , a binary operation  $+: I \times I \rightarrow I$ , and a distinguished element  $0 \in I$  with the following properties:

- (i) The operation  $+$  is commutative and associative, and has  $0$  as the identity element.
- (ii) The partial ordering  $\prec$  has  $0$  as the minimal element.
- (iii) If  $i, j \in I$  are such that  $i \prec j$ , then  $i + k \prec j + k$  for all  $k \in I$ .
- (iv) For every  $d \in I$ , the initial segment  $\{i \in I : i \prec d\}$  is finite.

A *finite downset* in  $I$  is a finite subset  $J$  of  $I$  with the property that  $j \in J$  whenever  $j \in I$  and  $j \prec i$  for some  $i \in J$ .

In this paper, we will only need the following three specific orderings (with  $k$  a standard positive integer):

- (i) The *degree ordering*, in which  $I = \mathbb{N}$  with the usual ordering, addition, and zero element.
- (ii) The *multidegree ordering*, in which  $I = \mathbb{N}^k$  with the usual addition and zero element, and with the product ordering, thus  $(i'_1, \dots, i'_k) \preceq (i_1, \dots, i_k)$  if  $i'_j \leq i_j$  for all  $1 \leq j \leq k$ .
- (iii) The *degree-rank ordering*, in which  $I$  is the sector  $\text{DR} := \{(d, r) \in \mathbb{N}^2 : 0 \leq r \leq d\}$  with the usual addition and zero element, and the lexicographical ordering, that is to say  $(d', r') \prec (d, r)$  if  $d' < d$  or if  $d' = d$  and  $r' < r$ .

It is easy to verify that each of these three explicit orderings obeys the abstract axioms in Definition 6.7. In the case of the degree or degree-rank orderings,  $I$  is totally ordered (for instance, the first few degree-ranks are  $(0, 0), (1, 0), (1, 1), (2, 0), (2, 1), (2, 2), (3, 0), \dots$ ), and so the only finite downsets are the initial segments. For the multidegree ordering, however, the initial segments are not the only finite downsets that can occur.

The one-dimensional notions of a filtration, nilsequence, nilcharacter, and symbol can be easily generalised to arbitrary orderings. We give the bare definitions here, and defer the more thorough treatment of these concepts to Appendix B and Appendix E. We will however remark that when  $I$  is the degree ordering, then all of the notions defined below simplify to the one-dimensional counterparts defined earlier.

**Definition 6.8** (Filtered group). Let  $I$  be an ordering and let  $G$  be a group. By an  *$I$ -filtration* on  $G$  we mean a collection  $G_I = (G_i)_{i \in I}$  of subgroups indexed by  $I$ , with the following properties:

- (i) (Nesting) If  $i, j \in I$  are such that  $i \prec j$ , then  $G_i \supseteq G_j$ .
- (ii) (Commutators) For every  $i, j \in I$ , we have  $[G_i, G_j] \subseteq G_{i+j}$ .

If  $d \in I$ , we say that  $G$  has *degree*  $\leq d$  if  $G_i$  is trivial whenever  $i \not\preceq d$ . More generally, if  $J$  is a downset in  $I$ , we say that  $G$  has *degree*  $\subseteq J$  if  $G_i$  is trivial whenever  $i \notin J$ .

Let us explicitly adapt the above abstract definitions to the three specific orderings mentioned earlier.

**Definition 6.9.** If  $(d_1, \dots, d_k) \in \mathbb{N}^k$ , we define a *nilpotent Lie group of multi-degree*  $\leq (d_1, \dots, d_k)$  to be a nilpotent  $I$ -filtered Lie group of degree  $\leq (d_1, \dots, d_k)$ , where  $I = \mathbb{N}^k$  is the multidegree ordering. Similarly, if  $J$  is a downset, define the notion of a nilpotent Lie group of multidegree  $\subseteq J$ .

If  $(d, r) \in \text{DR}$ , define a *nilpotent Lie group of degree-rank  $\leq (d, r)$*  to be a nilpotent DR-filtered Lie group  $G$  of degree  $\leq (d, r)$ , with the additional axioms  $G_{(0,0)} = G$  and  $G_{(d,0)} = G_{(d,1)}$  for all  $d \geq 1$ .

We define the notion of a filtered nilmanifold of multidegree  $\leq (d_1, \dots, d_k)$ , multidegree  $\subseteq J$ , or degree-rank  $\leq (d, r)$  similarly.

Note that the degree-rank filtration needs to obey some additional axioms, which are needed in order for the rank  $r$  to play a non-trivial role. As such, the unification here of degree, multidegree, and degree-rank, is not quite perfect; however this wrinkle is only of minor technical importance and should be largely ignored on a first reading.

**Example 6.10.** If  $G$  is a filtered nilpotent group of multidegree  $\leq (1, 1)$ , then the groups  $G_{(1,0)}$  and  $G_{(0,1)}$  must be abelian normal subgroups of  $G_{(0,0)}$ , and their commutator  $[G_{(1,0)}, G_{(0,1)}]$  must lie inside the group  $G_{(1,1)}$ , which is a central subgroup of  $G_{(0,0)}$ .

If  $G$  is a filtered nilpotent group of degree-rank  $\leq (d, d)$ , then  $(G_{(i,0)})_{i \geq 0}$  is a  $\mathbb{N}$ -filtration of degree  $\leq d$ . But if we reduce the rank  $r$  to be strictly less than  $d$ , then we obtain some additional relations between the  $G_{(i,0)}$  that do not come from the filtration property. For instance, if  $G$  has degree-rank  $\leq (3, 2)$ , then the group  $[G_{(1,0)}, [G_{(1,0)}, G_{(1,0)}]]$  must now be trivial; if  $G$  has degree-rank  $\leq (3, 1)$ , then the group  $[G_{(1,0)}, G_{(2,0)}]$  must also be trivial. More generally, if  $G$  has degree-rank  $\leq (d, r)$ , then any iterated commutator of  $g_{i_1}, \dots, g_{i_m}$  with  $g_j \in G_{(i_j,0)}$  for  $j = 1, \dots, m$  will be trivial whenever  $i_1 + \dots + i_m > d$ , or if  $i_1 + \dots + i_m = d$  and  $m > r$ .

**Example 6.11.** If  $(G_i)_{i \in \mathbb{N}}$  is an  $\mathbb{N}$ -filtration of  $G$  of degree  $\leq d$ , then  $(G_{|\vec{i}|})_{\vec{i} \in \mathbb{N}^k}$  is an  $\mathbb{N}^k$ -filtration of  $G$  of multidegree  $\subset \{\vec{i} \in \mathbb{N}^k : |\vec{i}| \leq d\}$ , where we recall the notational convention  $|(i_1, \dots, i_k)| = i_1 + \dots + i_k$ . Conversely, if  $J$  is a finite downset of  $\mathbb{N}^k$  and  $(G_{\vec{i}})_{\vec{i} \in \mathbb{N}^k}$  is a  $\mathbb{N}^k$ -filtration of  $G$  of multidegree  $\subset J$ , then

$$\left( \bigvee_{\vec{i}: |\vec{i}| \leq i} G_{\vec{i}} \right)_{i \in \mathbb{N}}$$

is easily verified (using Lemma 3.1) to be an  $\mathbb{N}$ -filtration of degree  $\leq \max_{\vec{i} \in J} |\vec{i}|$ , where  $\bigvee_{a \in A} G_a$  is the group generated by  $\bigcup_{a \in A} G_a$ . In particular, any multidegree  $\leq (d_1, \dots, d_k)$  filtration induces a degree  $\leq d_1 + \dots + d_k$  filtration.

In a similar spirit, every degree-rank  $\leq (d, r)$  filtration  $(G_{(d', r')})_{(d', r') \in \text{DR}}$  of a group  $G$  induces a degree  $\leq d$  filtration  $(G_{(i,0)})_{i \in \mathbb{N}}$ . In the converse direction, if  $(G_i)_{i \in \mathbb{N}}$  is a degree  $\leq d$  filtration of  $G$  with  $G = G_0$ , then we can create a degree-rank  $\leq (d, d)$  filtration  $(G_{(d', r')})_{(d', r') \in \text{DR}}$  by setting  $G_{(d', r')}$  to be the space generated by all the iterated commutators of  $g_{i_1}, \dots, g_{i_m}$  with  $g_j \in G_{(i_j,0)}$  for  $j = 1, \dots, m$  for which either  $i_1 + \dots + i_m > d'$ , or  $i_1 + \dots + i_m = d$  and  $m \geq \max(r', 1)$ ; this can easily be verified to indeed be a filtration, thanks to Lemma 3.1.

**Example 6.12.** Let  $d \geq 1$  be a standard integer. We can give the unit circle  $\mathbb{T}$  the structure of a degree-rank filtered nilmanifold of degree-rank  $\leq (d, 1)$  by setting  $G = \mathbb{R}$  and  $\Gamma = \mathbb{Z}$  with  $G_{(d', r')} := \mathbb{R}$  for  $(d', r') \leq (d, 1)$  and  $G_{(d', r')} := \{0\}$  otherwise. This is also the filtration obtained from the degree  $\leq d$  filtration (see Example 4.3) using the construction in Example 6.11.

**Example 6.13** (Products). If  $G_I$  and  $G'_I$  are  $I$ -filtrations on groups  $G, G'$  then we can give the product  $G \times G'$  an  $I$ -filtration in an obvious way by setting  $(G \times G')_i := G_i \times G'_i$ . The degree of  $G \times G'$  is the union of the degrees of  $G$  and  $G'$ . Similarly the product  $G_1/\Gamma_1 \times G_2/\Gamma_2$  of two  $I$ -filtered nilmanifolds is an  $I$ -filtered nilmanifold.

**Example 6.14** (Pushforward and pullback). Let  $\phi : G \rightarrow H$  be a homomorphism of groups. Then any  $I$ -filtration  $H_I = (H_i)_{i \in I}$  of  $H$  induces a *pullback  $I$ -filtration*  $\phi^* H_I := (\phi^{-1}(H_i))_{i \in I}$ . Similarly, any  $I$ -filtration  $G_I = (G_i)_{i \in I}$  on  $G$  induces a *pushforward  $I$ -filtration*  $\phi_* G_I := (\phi(G_i))_{i \in I}$  on  $H$ . In particular, if  $\Gamma$  is a subgroup of  $G$ , then we can pullback a filtration  $G_I = (G_i)_{i \in I}$  of  $G$  by the inclusion map  $\iota : \Gamma \hookrightarrow G$  to create the *restriction*  $\Gamma_I := (\Gamma_i)_{i \in I}$  of that filtration. It is a trivial matter to check that the subgroups of this filtration are given by  $\Gamma_i := \Gamma \cap G_i$ .

**Definition 6.15** (Filtered quotient space). A  *$I$ -filtered quotient space* is a quotient  $G/\Gamma$ , where  $G$  is an  $I$ -filtered group and  $\Gamma$  is a subgroup of  $G$  (with the induced filtration, see Example 6.14).

A  *$I$ -filtered homomorphism*  $\phi : G/\Gamma \rightarrow G'/\Gamma'$  between  $I$ -filtered quotient spaces is a group homomorphism  $\phi : G \rightarrow G'$  which maps  $\Gamma$  to  $\Gamma'$ , and also maps  $G_i$  to  $G'_i$  for all  $i \in I$ . Note that such a homomorphism descends to a map from  $G/\Gamma$  to  $G'/\Gamma'$ .

If  $G$  is a nilpotent  $I$ -filtered Lie group, and  $\Gamma$  is a discrete cocompact subgroup of  $G$  which is rational with respect to  $G_I$  (thus  $\Gamma_i := \Gamma \cap G_i$  is cocompact in  $G_i$  for each  $i \in I$ ), we call  $G/\Gamma = (G/\Gamma, G_I)$  an  *$I$ -filtered nilmanifold*. We say that  $G/\Gamma$  has degree  $\leq d$  or  $\subseteq J$  if  $G$  has degree  $\leq d$  or  $\subseteq J$ .

**Example 6.16** (Subnilmanifolds). Let  $G/\Gamma$  be an  $I$ -filtered nilmanifold of degree  $\subseteq J$ . If  $H$  is a rational subgroup of  $G$ , then  $H/(H \cap \Gamma)$  is also a filtered nilmanifold degree  $\subseteq J$  (using Example 6.14), with an inclusion homomorphism from  $H/(H \cap \Gamma)$  to  $G/\Gamma$ ; we refer to  $H/(H \cap \Gamma)$  as a *subnilmanifold* of  $G/\Gamma$ .

We isolate three important examples of a filtered group, in which  $G$  is the additive group  $\mathbb{Z}$  or  $\mathbb{Z}^k$ .

**Definition 6.17** (Basic filtrations). We define the following filtrations:

- The *degree filtration*  $\mathbb{Z}_{\mathbb{N}}^k$  on  $G = \mathbb{Z}^k$ , in which  $I = \mathbb{N}$  is the degree ordering and  $G_i = G$  for  $i \leq 1$  and  $G_i = \{0\}$  otherwise. In many cases  $k$  will equal 1 or 2.
- The *multidegree filtration*  $\mathbb{Z}_{\mathbb{N}^k}^k$  on  $G = \mathbb{Z}^k$ , in which  $I = \mathbb{N}^k$  is the multidegree ordering and  $G_{\vec{0}} = \mathbb{Z}^k$ ,  $G_{\vec{e}_i} = \langle \vec{e}_i \rangle$ ,  $i = 1, \dots, k$ , and  $G_{\vec{v}} = \{0\}$  otherwise, with  $e_1, \dots, e_k$  being the standard basis for  $\mathbb{Z}^k$ ;
- The *degree-rank filtration*  $\mathbb{Z}_{\text{DR}}$  on  $G = \mathbb{Z}$ , in which  $I = \text{DR}$  is the degree-rank ordering and  $G_{(0,0)} = G_{(1,0)} = \mathbb{Z}$  and  $G_{(d,r)} = \{0\}$  otherwise.

**Definition 6.18** (Polynomial map). Suppose that  $H$  and  $G$  are  $I$ -filtered groups with  $H = (H, +)$  abelian<sup>4</sup>. Then for any map  $g : H \rightarrow G$  we define the derivative

$$\partial_h g(n) := g(n+h)g(n)^{-1}. \quad (6.6)$$

We say that  $g : H \rightarrow G$  is *polynomial* if

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_{i_1 + \dots + i_m} \quad (6.7)$$

<sup>4</sup>This is not actually a necessary assumption; see Appendix B. However, in the main body of the paper we will only be concerned with polynomial maps on additive domains.

for all  $m \geq 0$ , all  $i_1, \dots, i_m \in I$  and all  $h_j \in H_{i_j}$  for  $j = 1, \dots, m$ , and for all  $n \in H_0$ .

We denote by  $\text{poly}(H_I \rightarrow G_I)$  the space of all polynomial maps from  $H_I$  to  $G_I$ . As usual, we use  $^*\text{poly}(H_I \rightarrow G_I)$  to denote the space of all limit polynomial maps from  $^*H_I$  to  $^*G_I$  (i.e. ultralimits of polynomial maps in  $\text{poly}(H_I \rightarrow G_I)$ ).

Many facts about these spaces (in some generality) are established in Appendix B where, in particular, a remarkable result essentially due to Lazard and Leibman [42, 43, 44] is established:  $\text{poly}(H_I \rightarrow G_I)$  is a group. The material in Appendix B is formulated in the general setting of abstract orderings  $I$  and for arbitrary (and possibly non-abelian) groups  $H_I$ , but for our applications we are only interested in the special case when  $H_I$  is  $\mathbb{Z}$  or  $\mathbb{Z}^k$  with the degree, multidegree, or degree-rank filtration as defined above.

Before moving on let us be quite explicit about what the notion of a polynomial map is in each of the three cases, since the definitions take a certain amount of unravelling.

- (Degree filtration) If  $H = \mathbb{Z}^k$  with the degree filtration  $\mathbb{Z}_{\mathbb{N}}^k$ , then  $\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$  consists of maps  $g : \mathbb{Z}^k \rightarrow G$  with the property that

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_m$$

for all  $m \geq 0$ ,  $h_1, \dots, h_m \in \mathbb{Z}$  and all  $n \in G_0$ . This space is precisely the same space as the one considered in [24, §6]. The space  $^*\text{poly}(\mathbb{Z}^k \rightarrow G_{\mathbb{N}})$  is defined similarly, except that  $g : ^*\mathbb{Z}^k \rightarrow ^*G$  is now a limit map, and all spaces such as  $\mathbb{Z}$  and  $G_m$  need to be replaced by their ultrapowers. (Similarly for the other two examples in this list.)

- (Multidegree filtration) If  $H = \mathbb{Z}^k$  with the multidegree filtration  $\mathbb{Z}_{\mathbb{N}^k}^k$ , then  $\text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow G_{\mathbb{N}^k})$  consists of maps  $g : \mathbb{Z}^k \rightarrow G$  with the property that

$$\partial_{\vec{e}_{i_1}} \dots \partial_{\vec{e}_{i_m}} g(\vec{n}) \in G_{\vec{e}_{i_1} + \dots + \vec{e}_{i_m}}$$

for all  $k \geq 0$ , all  $i_1, \dots, i_m$  and all  $\vec{n} \in \mathbb{Z}^k$ . To relate this space to the analogous spaces for the degree ordering, observe (using Example 6.11) that

$$\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow (G_i)_{i \in \mathbb{N}}) = \text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow (G_{|\vec{i}|})_{\vec{i} \in \mathbb{N}^k})$$

for any  $\mathbb{N}$ -filtration  $(G_i)_{i \in \mathbb{N}}$ , and conversely one has

$$\text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow (G_{\vec{i}})_{\vec{i} \in \mathbb{N}^k}) \subset \text{poly}\left(\mathbb{Z}_{\mathbb{N}}^k \rightarrow \left(\bigvee_{|\vec{i}|=i} G_{\vec{i}}\right)_{i \in \mathbb{N}}\right)$$

for any  $\mathbb{N}^k$ -filtration  $(G_{\vec{i}})_{\vec{i} \in \mathbb{N}^k}$ . This is of course related to the obvious fact that a polynomial of multidegree  $\leq (d_1, \dots, d_k)$  is automatically of degree  $\leq d_1 + \dots + d_k$ .

- (Degree-rank filtration) If  $H = \mathbb{Z}$  with the degree-rank filtration  $\mathbb{Z}_{\text{DR}}$ ,  $\text{poly}(\mathbb{Z}_{\text{DR}} \rightarrow G_{\text{DR}})$  consists of maps  $g : \mathbb{Z} \rightarrow G$  with the property that

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_{(m,0)}$$

whenever  $m \geq 0$ ,  $h_1, \dots, h_m \in \mathbb{Z}$  and  $n \in G_0$ . We observe (using Example 6.11) the obvious equality

$$\text{poly}(\mathbb{Z}_{\text{DR}} \rightarrow (G_{(d,r)})_{(d,r) \in \text{DR}}) = \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G_{(i,0)})_{i \in \mathbb{N}}) \quad (6.8)$$

for any DR-filtration  $(G_{(d,r)})_{(d,r) \in \text{DR}}$ . Thus, a degree-rank filtration  $G_{\text{DR}}$  on  $G$  does not change the notion of a polynomial sequence, but instead gives some finer information on the group  $G$  (and in particular, it indicates that certain iterated commutators of the  $G_{(d,r)}$  vanish, which is information that cannot be discerned just from the knowledge that  $(G_{(i,0)})_{i \in \mathbb{N}}$  is a  $\mathbb{N}$ -filtration).

**Definition 6.19** (Nilsequences and nilcharacters). Let  $I$  be an ordering, and let  $J$  be a finite downset in  $I$ . Let  $H$  be an abelian  $I$ -filtered group. A (polynomial) nilsequence of degree  $\subset J$  is any function of the form

$$\chi(n) = F(g(n)^*\Gamma),$$

where

- $G/\Gamma = (G/\Gamma, G_I)$  is an  $I$ -filtered nilpotent manifold of degree  $\subset J$ ;
- $g \in {}^*\text{poly}(H_I \rightarrow G_I)$  is a limit polynomial map from  ${}^*H_I$  to  ${}^*G_I$ ; and
- $F \in \text{Lip}({}^*(G/\Gamma) \rightarrow \overline{C}^\omega)$ .

The space of all such nilsequences will be denoted  $\text{Nil}^{\subset J}({}^*H)$ . We define the notion of a nilsequence of degree  $\leq d$  for some  $d \in I$ , and the space  $\text{Nil}^{\leq d}({}^*H)$ , similarly. If  $\Omega$  is a limit subset of  ${}^*H$ , the restriction of the nilsequences in  $\text{Nil}^{\subset J}({}^*H)$  to  $\Omega$  will be denoted  $\text{Nil}^{\subset J}(\Omega)$ , and we define  $\text{Nil}^{\leq d}(\Omega)$  similarly.

We refer to the map  $n \mapsto g(n)^*\Gamma$  as a *limit polynomial orbit* in  $G/\Gamma$ , and denote the space of such orbits as  ${}^*\text{poly}(H_I \rightarrow (G/\Gamma)_I)$ .

Suppose that  $d \in I$ . Then  $\chi$  is said to be a *degree  $d$  nilcharacter* if  $\chi$  is a degree  $\leq d$  nilsequence with the following additional properties:

- $F \in \text{Lip}({}^*(G/\Gamma) \rightarrow \overline{S}^\omega)$  (thus  $|F| = 1$ ) and
- $F(g_d x) = e(\eta(g_d))F(x)$  for all  $x \in G/\Gamma$  and  $g_d \in G_d$ , where  $\eta : G_d \rightarrow \mathbb{R}$  is a continuous standard homomorphism which maps  $\Gamma_d$  to the integers. We call  $\eta$  the *vertical frequency* of  $F$ .

The space of all degree  $d$  nilcharacters on  ${}^*H$  will be denoted  $\Xi^d({}^*H)$ . If  $\Omega$  is a limit subset of  ${}^*H$ , the restriction of the nilcharacters in  $\Xi^d({}^*H)$  to  $\Omega$  will be denoted  $\Xi^d(\Omega)$ .

With the multidegree ordering, a degree  $(d_1, \dots, d_k)$  nilcharacter will be referred to as a multidegree  $(d_1, \dots, d_k)$  nilcharacter, and the space of such characters on  $\Omega$  denoted  $\Xi_{\text{Multi}}^{(d_1, \dots, d_k)}(\Omega)$ ; we similarly write  $\text{Nil}^{\subset J}(\Omega)$  or  $\text{Nil}^{\leq (d_1, \dots, d_k)}(\Omega)$  as  $\text{Nil}_{\text{Multi}}^{\subset J}(\Omega)$  or  $\text{Nil}^{\leq (d_1, \dots, d_k)}(\Omega)$  for emphasis.

Similarly, with the degree-rank ordering, and assuming  $G/\Gamma$  is a filtered nilmanifold of degree-rank  $\leq (d, r)$  (so in particular, we enforce the axioms  $G_{(0,0)} = G$  and  $G_{(d,0)} = G_{(d,1)}$ ), a degree  $(d, r)$  nilcharacter will be referred to as a degree-rank  $(d, r)$  nilcharacter. The space of nilcharacters on  $\Omega$  of degree-rank  $(d, r)$  will be denoted  $\Xi_{\text{DR}}^{(d,r)}(\Omega)$  (note that this is distinct from the space  $\Xi_{\text{Multi}}^{(d_1, d_2)}(\Omega)$  of two-dimensional nilcharacters of multidegree  $(d_1, d_2)$ ), and the nilsequences on  $\Omega$  of degree-rank  $\leq (d, r)$  will similarly be denoted  $\text{Nil}_{\text{DR}}^{\leq (d,r)}(\Omega)$ .

**Example 6.20.** Let  $J \subset \mathbb{N}^k$  be a finite downset. Then any sequence of the form

$$(n_1, \dots, n_k) \mapsto F \left( \sum_{(i_1, \dots, i_k) \in J} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k} \bmod 1 \right),$$

where  $\alpha_{i_1, \dots, i_k} \in {}^*\mathbb{R}$  and  $F \in \text{Lip}({}^*\mathbb{T} \rightarrow \overline{\mathbb{C}}^\omega)$ , is a nilsequence on  $\mathbb{Z}^k$  of multidegree  $\subseteq J$ , as can easily be seen by giving  $G := \mathbb{R}$  the  $\mathbb{Z}^k$ -filtration  $G_i := \mathbb{R}$  for  $i \in J$  and  $G_i := \{0\}$  otherwise, and setting  $\Gamma := \mathbb{Z}$  and  $g \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow \mathbb{R})$  to be the limit polynomial  $n \mapsto \sum_{(i_1, \dots, i_k) \in J} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k}$ .

For similar reasons, any sequence of the form

$$(n_1, \dots, n_k) \mapsto e \left( \sum_{(i_1, \dots, i_k) \in \mathbb{N}^k: i_1 + \dots + i_k \leq d} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k} \bmod 1 \right),$$

is a degree  $d$  nilcharacter on  $\mathbb{Z}^k$  of degree  $d$ , and any sequence of the form

$$(n_1, \dots, n_k) \mapsto e \left( \sum_{(i_1, \dots, i_k) \in \mathbb{N}^k: i_j \leq d_j \text{ for } j=1, \dots, k} \alpha_{i_1, \dots, i_k} n_1^{i_1} \dots n_k^{i_k} \bmod 1 \right),$$

is a multidegree  $(d_1, \dots, d_k)$  nilcharacter on  $\mathbb{Z}^k$ .

**Example 6.21.** Any degree 2 nilsequence of magnitude 1 is automatically a degree-rank  $\leq (3, 0)$  nilcharacter, since every degree  $\leq 2$  nilmanifold is automatically a degree-rank  $\leq (2, 2)$  nilmanifold, which can then converted trivially to a degree-rank  $\leq (3, 0)$  nilmanifold (with a trivial group  $G_{(3,0)}$ ). Thus for instance for  $\alpha, \beta \in \mathbb{R}$ ,

$$n \mapsto e(\{\alpha n\}\beta n)$$

is nearly a degree-rank  $(3, 0)$  nilcharacter, and becomes a genuine degree-rank  $(3, 0)$  nilcharacter after vector-valued smoothing.

If  $\alpha \in {}^*\mathbb{R}$ , then the sequence

$$n \mapsto e(\alpha n^3)$$

is a degree-rank  $(3, 1)$  nilcharacter. Indeed, we can give  $G = \mathbb{R}$  a degree-rank  $\leq (3, 1)$  filtration  $G_{\text{DR}}$  by setting  $G_{(d,r)} := \mathbb{R}$  for  $(d, r) \leq (3, 1)$ , and  $G_{(d,r)} := \{0\}$  otherwise.

Next, if  $\alpha, \beta \in {}^*\mathbb{R}$ , then the sequence

$$n \mapsto e(\{\alpha n^2\}\beta n) \tag{6.9}$$

is *nearly* a degree-rank  $(3, 2)$  nilcharacter (and becomes a genuinely so after vector-valued smoothing). To see this, let  $G$  be the Heisenberg nilpotent group (6.1), which we give the following degree-rank filtration:

$$\begin{aligned} G_{(0,0)} &= G_{(1,0)} = G_{(1,1)} := G \\ G_{(2,0)} &= G_{(2,1)} := \langle e_1, [e_1, e_2] \rangle_{\mathbb{R}} = \{e_1^{t_1} [e_1, e_2]^{t_{12}} : t_1, t_{12} \in \mathbb{R}\} \\ G_{(2,2)} &= G_{(3,0)} = G_{(3,1)} = G_{(3,2)} := \langle [e_1, e_2] \rangle_{\mathbb{R}} = \{[e_1, e_2]^{t_{12}} : t_{12} \in \mathbb{R}\} \\ G_{(d,r)} &:= \{\text{id}\} \text{ for all other } (d, r) \in \text{DR}. \end{aligned}$$

One easily verifies that this is a degree-rank  $\leq (3, 2)$  filtration. If we then set  $g : {}^*\mathbb{Z} \rightarrow {}^*G$  to be the limit sequence  $g(n) := e_2^{\beta n} e_1^{\alpha n^2}$ , one easily verifies that  $g$  is a limit polynomial with respect to this degree-rank filtration. If one then lets  $F$  be the piecewise Lipschitz function (6.2), then we see that

$$F(g(n){}^*\Gamma) = e(\{\alpha n^2\}\beta n)$$

and so we see that  $n \mapsto e(\{\alpha n^2\}\beta n)$  is a indeed piecewise degree-rank  $(3, 2)$  nilcharacter.

A similar argument (using the free 3-step nilpotent manifold on three generators, which has degree  $\leq 3$  and hence degree-rank  $\leq (3, 3)$ ) shows that

$$n \mapsto e(\{\{\alpha n\}\beta n\}\gamma n)$$

is nearly a degree-rank  $(3, 3)$  nilcharacter, and becomes a genuine degree-rank  $(3, 3)$  nilcharacter after applying vector-valued smoothing; see [28, Appendix E] for the relevant calculations.

These examples should help illustrate the heuristic that a degree-rank  $(d, r)$  nilcharacter is built up using (suitable vector-valued smoothings of) bracket monomials which either have degree less than  $d$ , or have degree exactly  $d$  and involve at most  $r - 1$  applications of the fractional part operation.

We observe (using Example 6.11) the following obvious inclusions:

- (i) A multidegree  $\leq (d_1, \dots, d_k)$  nilsequence on  $\mathbb{Z}^k$  is automatically a degree  $\leq d_1 + \dots + d_k$  nilsequence.
- (ii) A multidegree  $(d_1, \dots, d_k)$  nilcharacter on  $\mathbb{Z}^k$  is automatically a degree  $d_1 + \dots + d_k$  nilcharacter.
- (iii) A multidegree  $(d_1, \dots, d_{k-1}, 0)$  nilsequence on  $\mathbb{Z}^k$  is constant in the  $n_k$  variable, and descends to a multidegree  $(d_1, \dots, d_{k-1})$  nilsequence on  $\mathbb{Z}^{k-1}$ .
- (iv) A degree-rank  $\leq (d, r)$  nilsequence on  $\mathbb{Z}$  is automatically a degree  $\leq d$  nilsequence.
- (v) A degree  $\leq d$  nilsequence on  $\mathbb{Z}$  is automatically a degree-rank  $\leq (d, d)$  nilsequence.
- (vi) A degree  $d$  nilcharacter on  $\mathbb{Z}$  is automatically a degree-rank  $\leq (d, d)$  nilcharacter.

It is not quite true, though, that a degree-rank  $(d, r)$  nilcharacter is a degree  $d$  nilcharacter if  $r > 1$ , because the former need not exhibit vertical frequency behaviour for degree-ranks  $(d, r')$  with  $r' < r$ .

**Definition 6.22** (Equivalence and symbols). Let  $H$  be an  $I$ -filtered group, let  $d \in I$ , and let  $\Omega$  be a limit subset of  ${}^*H$ . Two nilcharacters  $\chi, \chi' \in \Xi^d(\Omega)$  are said to be *equivalent* if  $\chi \otimes \overline{\chi'}$  is a nilsequence of degree strictly less than  $d$ . Write  $[\chi]_{\text{Symb}^d(\Omega)}$  for the equivalence class of  $\chi$  with respect to this relation; this we shall refer to as the *symbol* of  $\chi$ . Write  $\text{Symb}^d(\Omega)$  for the space of all such equivalence classes.

We write  $\text{Symb}_{\text{Multi}}^{(d_1, \dots, d_k)}(\Omega)$  for the symbols of nilcharacters  $\chi \in \Xi_{\text{Multi}}^{(d_1, \dots, d_k)}(\Omega)$  of multidegree  $(d_1, \dots, d_k)$ , and  $\text{Symb}_{\text{DR}}^{(d, r)}(\Omega)$  for the symbols of nilcharacters  $\chi \in \Xi_{\text{DR}}^{(d, r)}(\Omega)$  of degree-rank  $(d, r)$ . The basic properties of such symbols are set out in Appendix E.

## 7. A MORE DETAILED OUTLINE OF THE ARGUMENT

Now that we have set up the notation to describe nilcharacters and their symbols, we are ready to give a high-level proof of Conjecture 5.5 (and hence Theorem 1.3), contingent on some key sub-theorems which will be proven in later sections. This corresponds to the realisation of points (i), (ii) and (ix) from the overview in §2.

As the cases  $s = 1, 2$  of this conjecture are already known, we assume that  $s \geq 3$ . We also assume inductively that the claim has already been proven for smaller values of  $s$ . Henceforth  $s$  is fixed.

Let  $f \in L^\infty[N]$  be such that

$$\|f\|_{U^{s+1}[N]} \gg 1. \quad (7.1)$$

Define  $f$  to be zero outside of  $[N]$ . Raising (7.1) to the power  $2^{s+1}$ , we see that

$$\mathbb{E}_{h \in [[N]]} \|\Delta_h f\|_{U^s[N]}^{2^s} \gg 1$$

and thus

$$\|\Delta_h f\|_{U^s[N]} \gg 1$$

for all  $h$  in a dense subset  $H$  of  $[[N]]$ . Applying the inductive hypothesis, we thus see that  $\Delta_h f$  is  $(s-1)$ -biased for all  $h \in H$ .

By definition, we now know that  $\Delta_h f$  correlates with a nilsequence of degree  $(s-1)$ . By Lemma 6.5, we see that for each  $h \in H$ ,  $\Delta_h f$  correlates with a nilcharacter  $\chi_h \in \Xi^{s-1}([N])$ . It is not hard to see that the space of such nilcharacters is a  $\sigma$ -limit set (see Definition A.10), so by Lemma A.12 we can ensure that  $\chi_h$  depends in a limit fashion on  $h$ .

The aim at this point is to obtain, in several stages, information about the dependence of  $\chi_h$  on  $h$ . A key milestone in this analysis is a *linearisation* of  $\chi_h$  on  $h$ . In the case  $s=2$ , treated in [16, 21], the  $\chi_h(n)$  were essentially just linear phases  $e(\xi_h n)$ , and the outcome of the linearisation analysis was that the frequencies  $\xi_h$  may be assumed to vary in a bracket-linear fashion with  $h$ . In the case  $s=3$  (treated in [28] but also dealt with in our present work), a model special case occurs when  $\chi_h(n) \approx e(\{\alpha_h n\}\beta_h n)$  (interpreting  $\approx$  loosely). The outcome of the linearisation analysis in that case was that at most one of  $\alpha_h, \beta_h$  really depends on  $h$ , and furthermore that this dependence on  $h$  is bracket-linear in nature.

Now we formally set out the general case of this linearisation process.

**Theorem 7.1** (Linearisation). *Let  $f \in L^\infty[N]$ , let  $H$  be a dense subset of  $[[N]]$ , and let  $(\chi_h)_{h \in H}$  be a family of nilcharacters in  $\Xi^{s-1}([N])$  depending in a limit fashion on  $h$ , such that  $\Delta_h f$  correlates with  $\chi_h$  for all  $h \in H$ . Then there exists a multidegree  $(1, s-1)$ -nilcharacter  $\chi \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$  such that  $\Delta_h f$   $(s-2)$ -correlates with  $\chi(h, \cdot)$  for many  $h \in H$ .*

This statement represents the outcome of points (iii) to (vii) of the outline in §2 and must therefore address the following points:

- For some suitable notion of “frequency”, the symbol of  $\chi_h(n)$  contains only one frequency that genuinely depends on  $h$ ;
- That frequency depends on  $h$  in a bracket-linear manner;
- Once this is known, it follows that, for many  $h$ ,  $\Delta_h f$   $(s-2)$ -correlates with  $\chi(h, n)$ , where  $\chi$  is a certain 2-variable nilsequence.

These three tasks are, in fact, established together and in an incremental fashion. The nilcharacter  $\chi_h(n)$  is gradually replaced by objects of the form  $\chi'(h, n) \otimes \chi'_h(n)$  where  $\chi'(h, n)$  is a 2-dimensional nilcharacter of multidegree  $(1, s-1)$  and, at each stage, the nilcharacter  $\chi'_h(n)$  (which has so far not been shown to vary in any nice way with  $h$ ) is “simpler” than  $\chi_h(n)$ . The notion of *simpler* in this context is measured by the degree-rank filtration, a concept that was introduced in the previous section. Thus the result of a single pass over the three points listed above is the following subclaim.

**Theorem 7.2** (Linearisation, inductive step). *Let  $1 \leq r_* \leq s-1$ , let  $f \in L^\infty[N]$ , let  $H$  be a dense subset of  $[[N]]$ , let  $\chi \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$ , let  $(\chi_h)_{h \in H}$  be a family of*



nilcharacters of degree-rank  $(s-1, r_*)$  depending in a limit fashion on  $h$ , such that  $\Delta_h f$   $(s-2)$ -correlates with  $\chi(h, \cdot) \otimes \chi_h$  for all  $h \in H$ . Then there exists a dense subset  $H'$  of  $H$ , a multidegree  $(1, s-1)$ -nilcharacter  $\chi' \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$  and a family  $(\chi'_h)_{h \in H'}$  of nilcharacters of degree-rank  $(s-1, r_*-1)$  depending in a limit fashion on  $h$ , such that  $\Delta_h f$   $(s-2)$ -correlates with  $\chi'(h, \cdot) \otimes \chi'_h$  for all  $h \in H'$ .

Theorem 7.1 follows easily by inductive use of this statement, starting with  $r_*$  equal to  $s-1$  and using Theorem 7.2 iteratively to decrease  $r_*$  all the way to zero.

To prove Theorem 7.2, we follow steps (iii) to (vii) in the outline quite closely. The first step, which is the realisation of (iii), is a Gowers-style Cauchy-Schwarz inequality to eliminate the function  $f$  as well as the 2-dimensional nilcharacter  $\chi(h, n)$  and therefore obtain a statement concerning only the (so far) unstructured-in- $h$  object  $\chi_h(n)$ . Here is a precise statement of the outcome of this procedure; the proof of this proposition is the main business of §8.

**Proposition 7.3** (Gowers Cauchy-Schwarz argument). *Let  $f, H, \chi, (\chi_h)_{h \in H}$  be as in Theorem 7.2. Then the sequence*

$$n \mapsto \chi_{h_1}(n) \otimes \chi_{h_2}(n + h_1 - h_4) \otimes \overline{\chi_{h_3}(n)} \otimes \overline{\chi_{h_4}(n + h_1 - h_4)} \quad (7.2)$$

*is  $(s-2)$ -biased for many additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ .*

With this in hand, we reach the most complicated part of the argument. This is the use of Proposition 7.3 to study the “frequencies” of the nilcharacters  $\chi_h$  and the way they depend on  $h$ . Roughly speaking, the aim is to interpret the tensor product (7.2) as a nilsequence itself (depending on  $h_1, h_2, h_3, h_4$ ) and use results from [24] to analyse its equidistribution and bias properties.

To make proper sense of this one must first find a suitable “representation” of the  $\chi_h(n)$  in which the frequencies are either independent of  $h$ , depend in a bracket-linear fashion on  $h$ , or are appropriately *dissociated* in  $h$ , in the sense that the frequencies associated to (7.2) are “linearly independent” for most additive quadruples  $h_1 + h_2 = h_3 + h_4$ . This task is one of the more technical part of the papers and is performed in in §10; it incorporates the additive combinatorial step (vi) of the outline from §2. The precise statement of what we prove is Lemma 10.10, the “sunflower decomposition”.

The representation of the  $\chi_h$  (and hence of (7.2)) involves constructing a suitable polynomial orbit on something resembling a free nilpotent Lie group  $\tilde{G}$ ; this device also featured in [28, §5]. Once this is done, one applies the results from [24] to examine the orbit of this polynomial sequence on the corresponding nilmanifold  $\tilde{G}/\tilde{\Gamma}$ . The results of [24] assert (roughly speaking) that this orbit is close to the uniform measure on a subnilmanifold  $H\tilde{\Gamma}/\tilde{\Gamma}$ , where  $H \leq \tilde{G}$  is some closed subgroup. In §11, we then crucially apply a commutator argument of Furstenberg and Weiss that exploits some equidistribution information on projections of  $H$  to say something about this group  $H$ . The upshot of this critical phase of the argument is that the  $h$ -dependence of the frequencies of  $\chi_h$  cannot be dissociated in nature, and must instead be completely bracket-linear; the precise statement here is Theorem 11.1.

At this point in the argument, we have basically shown that the top-order behaviour (in the degree-rank order) of the nilcharacters  $\chi_h(n)$  is bracket-linear in  $h$ . To complete the proof of Theorem 7.2 (and hence of Theorem 7.1) it remains to carry out part (vii) of the outline, that is to say to interpret this bracket-linear

part of  $\chi_h(n)$  as a multidegree  $(1, s-1)$  nilcharacter  $\chi'(h, n)$ . This is the first part of the argument where some sort of “degree  $s$  nil-object” is actually constructed, and is thus a key milestone in the inductive derivation of  $\text{GI}(s)$  from  $\text{GI}(s-1)$ . As remarked previously, our construction here is a little more conceptual (and abstractly algebraic) than in previous works, which have been somewhat *ad hoc*. The construction is given in §12. At the end of that section we wrap up the proof of Theorem 7.1: by this point, all the hard work has been done.

With Theorem 7.1 in hand, we have completed the first seven steps of the outline. The only remaining substantial step is step (viii), the symmetry argument. Here is a formal statement of it:

**Theorem 7.4** (Symmetrisation). *Let  $f \in L^\infty[N]$ , let  $H$  be a dense subset of  $[[N]]$ , and let  $\chi \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$  be such that  $\Delta_h f < s-2$ -correlates with  $\chi(h, \cdot)$  for all  $h \in H$ . Then there exists a nilcharacter  $\Theta \in \Xi^s(*\mathbb{Z})$  (with the degree filtration) and a nilsequence  $\Psi \in \text{Nil}_{\text{Multi}}^{C, J}(*\mathbb{Z}^2)$ , with  $J \subset \mathbb{N}^2$  given by the downset*

$$J := \{(i, j) \in \mathbb{N}^2 : i + j \leq s-1\} \cup \{(i, s-i) : 2 \leq i \leq s\}, \quad (7.3)$$

*such that  $\chi(h, n)$  is a bounded linear combination of  $\Theta(n+h) \otimes \overline{\Theta(n)} \otimes \Psi(h, n)$ .*

The proof is given in §13. Informally, this theorem asserts that the multidimensional degree  $(1, s-1)$  nilcharacter  $\chi(h, n)$  can be expressed as a derivative  $\Theta(n+h) \otimes \overline{\Theta(n)}$  of a degree  $s$  nilcharacter  $\Theta$ , modulo “lower order terms”, which in this context means multidimensional nilsequences  $\Psi(h, n)$  that either have total degree  $\leq s-1$ , or are of degree at most  $s-2$  in the  $n$  variable.

The remaining task for this section is to show how to complete the proof of Conjecture 5.3 (and Theorem 1.3) from this point. From the discussion at the beginning of this section, we have already arrived at a situation in which the given function  $f \in L^\infty[N]$  has the property that  $\Delta_h f$  correlates with  $\chi_h$  for all  $h$  in a dense subset  $H$  of  $[[N]]$ , where  $(\chi_h)_{h \in H}$  be a family of nilcharacters in  $\Xi^{s-1}([N])$  depending in a limit fashion on  $h$ .

From Theorem 7.1 and Theorem 7.4 we see that for many  $h \in [[N]]$ ,  $\Delta_h f \leq s-2$ -correlates with the sequence

$$n \mapsto \Theta(n+h) \otimes \overline{\Theta(n)} \otimes \Psi(h, n).$$

The next step is to break up  $J$  and  $\Psi$  into simpler components, and our tool for this purpose shall be Lemma E.4. Applying this lemma for  $\varepsilon$  sufficiently small, followed by the pigeonhole principle, one can thus find scalar-valued nilsequences  $\psi, \psi'$  on  $*\mathbb{Z}^2$  (with the multidegree filtration) of multidegree

$$\subset \{(i, 0) \in \mathbb{N}^2 : i \leq s-1\}$$

and

$$\subset \{(i, j) \in \mathbb{N}^2 : i \leq s-2; i+j \leq s\}$$

respectively, such that for many  $h \in [[N]]$ ,  $\Delta_h f \leq (s-2)$ -correlates with

$$n \mapsto \Theta(n+h) \otimes \overline{\Theta(n)} \psi(h, n) \psi'(h, n).$$

For fixed  $h$ , the nilsequence  $\psi'(h, n)$  has degree  $\leq s-2$  and can thus be ignored. Also,  $\psi(h, n) = \psi(n)$  is of multidegree  $\leq (s-1, 0)$  and is thus independent of  $h$ , with  $n \mapsto \psi(n)$  being a degree  $\leq s-1$  nilsequence. Thus, for many  $h \in [[N]]$ ,  $\Delta_h f \leq s-2$ -correlates with

$$n \mapsto \Theta(n+h) \otimes \overline{\Theta(n)} \psi(n).$$

Applying the pigeonhole principle again, we can thus find scalar nilsequences  $\theta, \theta' \in \text{Nil}^{\leq s}(*\mathbb{Z})$  such that for many  $h \in [[N]]$ ,  $\Delta_h f \leq (s-2)$ -correlates with

$$n \mapsto \theta(n+h)\theta'(n)$$

(indeed one takes  $\theta, \theta'$  to be coefficients of  $\Theta$  and  $\overline{\Theta}\psi$  respectively). Applying the converse to GI( $s$ ) (Proposition 5.6), we conclude

$$\|f\overline{\theta}(\cdot+h)\overline{f\theta'}(\cdot)\|_{U^{s-1}[N]} \gg 1$$

for many  $h \in H$ . Averaging over  $h$  (using Corollary A.6 to obtain the required uniformity), we conclude that

$$\mathbb{E}_{h \in [[N]]} \|f\overline{\theta}(\cdot+h)\overline{f\theta'}(\cdot)\|_{U^{s-1}[N]}^{2^{s-1}} \gg 1.$$

Applying the Cauchy-Schwarz-Gowers inequality (see e.g. [57, Equation (11.6)]) we conclude that

$$\|f\overline{\theta}\|_{U^s[N]} \gg 1$$

and hence by the inductive hypothesis (Conjecture 5.5 for  $s-1$ ),  $f\overline{\theta}$  is  $\leq (s-1)$ -biased. Since  $\theta$  is a degree  $\leq s$  nilsequence, we conclude that  $f$  is  $\leq s$ -biased, as required. This concludes the proof of Conjecture 5.5, Conjecture 5.3, and hence Theorem 1.3, contingent on Theorem 7.1 and Theorem 7.4.

## 8. A VARIANT OF GOWERS'S CAUCHY-SCHWARZ ARGUMENT

The aim of this section is prove Proposition 7.3. Thus, we have standard integers  $1 \leq r_* \leq s-1$ , a function  $f \in L^\infty[N]$ , a dense subset  $H$  of  $[[N]]$ , a two-dimensional nilcharacter  $\chi \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$  of multidegree  $(1, s-1)$ , and a family  $(\chi_h)_{h \in H}$  of nilcharacters of degree-rank  $(s-1, r_*)$  depending in a limit fashion on  $h$ . We are given that  $\Delta_h f$   $(s-2)$ -correlates with  $\chi(h, \cdot) \otimes \chi_h$  for all  $h \in H$ . Our objective is to show that, for many additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ , the expression

$$n \mapsto \chi_{h_1}(n) \otimes \chi_{h_2}(n+h_1-h_4) \otimes \overline{\chi_{h_3}(n)} \otimes \overline{\chi_{h_4}(n+h_1-h_4)} \quad (8.1)$$

(where we extend the  $\chi_h$  by zero outside of  $[N]$ ) is  $(s-2)$ -biased.

The strategy, following the work of Gowers [16], is to start with the  $\leq s-2$ -correlation between  $\Delta_h f$  and  $\chi(h, \cdot)\chi_h$  and then apply the Cauchy-Schwarz inequality repeatedly to eliminate all terms involving  $f$ ,  $\chi(h, \cdot)$ , finally arriving at a correlation statement that only involves  $\chi_h$  (and lower order terms).

Unfortunately, there is a technical issue that prevents one from doing this directly, namely that the behaviour of  $\chi(h, \cdot)$  in  $h$  is not quite linear enough to ensure that these terms are completely eliminated by a Cauchy-Schwarz procedure. In order to overcome this issue, one must first prepare  $\chi$  into a better form, as follows. We need the following technical notion (which will not be used outside of this section):

**Definition 8.1.** A *linearised*  $(1, s-1)$ -function is a limit function  $\chi : (h, n) \rightarrow \overline{\mathbb{C}}^\omega$  which has a factorisation

$$\chi(h, n) = c(n)^h \psi(n) \quad (8.2)$$

where  $\psi \in L^\infty(\mathbb{Z} \rightarrow \overline{\mathbb{C}}^\omega)$  and  $c \in L^\infty(\mathbb{Z} \rightarrow S^1)$  are such that, for every  $h, l \in \mathbb{Z}$ , the sequence

$$n \mapsto c(n-l)^h \overline{c(n)}^h$$

is a degree  $\leq s-2$  nilsequence.

**Remark 8.2.** Heuristically, one should think of a linearised  $(1, s-1)$ -function as (a vector-valued smoothing of) a function of the form

$$(h, n) \mapsto e(P(n) + hQ(n))$$

where  $P, Q$  are bracket polynomials of degree  $s-1$ ; for instance,

$$(h, n) \mapsto e(\{\alpha n\}\beta n + \{\gamma n\}\delta nh)$$

is morally a linearised  $(1, 2)$  function. This should be compared with more general multidegree  $(1, 2)$  nilcharacters, such as

$$(h, n) \mapsto e(\{\{\alpha h\}\beta n\}\gamma n)$$

which are not quite linear in  $h$  because the dependence on  $h$  is buried inside one or more fractional part operations. Intuitively, the point is that one can use the laws of bracket algebra (such as (6.4)) to move the  $h$  outside of all the fractional part expressions (modulo lower order terms). While one can indeed develop enough of the machinery of bracket calculus to realise this intuition concretely, we will instead proceed by the more abstract machinery of nilmanifolds in order to avoid having to set up the bracket calculus.

The key preparation for this is the following.

**Proposition 8.3.** *Let  $\chi \in \Xi_{\text{Multi}}^{(1, s-1)}(*\mathbb{Z}^2)$  be a two-dimensional nilcharacter of multidegree  $(1, s-1)$ , and let  $\varepsilon > 0$  be standard. Then one can approximate  $\chi$  to within  $\varepsilon$  in the uniform norm by a bounded linear combination of linearised  $(1, s-1)$ -functions.*

*Proof.* From Definition 6.1, we can express

$$\chi(h, n) = F(g(h, n)^*\Gamma)$$

where  $G/\Gamma$  is a  $\mathbb{N}^2$ -filtered nilmanifold of multidegree  $\leq (1, s-1)$ ,  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}^2}^2 \rightarrow G_{\mathbb{N}^2})$  (with  $\mathbb{Z}^2$  being given the multidegree filtration  $\mathbb{Z}_{\mathbb{N}^2}^2$ ), and  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{S^\omega})$  has a vertical frequency  $\eta : G_{(1, s-1)} \rightarrow \mathbb{R}$ .

We consider the quotient map  $\pi : G/\Gamma \rightarrow G/(G_{(1,0)}\Gamma)$  from  $G/\Gamma$  onto the nilmanifold  $G/(G_{(1,0)}\Gamma)$ , which can be viewed as an  $\mathbb{N}$ -filtered nilmanifold of degree  $\leq s-1$  (where we  $\mathbb{N}$ -filter  $G/G_{(1,0)}$  using the subgroups  $G_{(0,i)}G_{(1,0)}/G_{(1,0)}$ ). The fibers of this map are isomorphic to  $T := G_{(1,0)}/\Gamma_{(1,0)}$ . Observe that  $G_{(1,0)}$  is abelian, and so  $T$  is a torus; thus  $G/\Gamma$  is a torus bundle over  $G/(G_{(1,0)}\Gamma)$  with structure group  $T$ . The idea is to perform Fourier analysis on this large torus  $T$ , as opposed to the smaller torus  $G_{(1, s-1)}/\Gamma_{(1, s-1)}$ , to improve the behaviour of the nilcharacter  $\chi$ .

We pick a metric on the base nilmanifold  $G/(G_{(1,0)}\Gamma)$  and a small standard radius  $\delta > 0$ , and form a smooth partition of unity  $1 = \sum_{k=1}^K \varphi_k$  on  $G/(G_{(1,0)}\Gamma)$ , where each  $\varphi_k \in \text{Lip}(G/(G_{(1,0)}\Gamma) \rightarrow \mathbb{C})$  is supported on an open ball  $B_k$  of radius  $r$ . This induces a partition  $\chi = \sum_{k=1}^K \tilde{\chi}_k$ , where

$$\tilde{\chi}_k(h, n) = F(g(h, n)^*\Gamma)\varphi_k(\pi(g(h, n)^*\Gamma)).$$

Now fix one of the  $k$ . Then we have

$$\tilde{\chi}_k(h, n) = \tilde{F}_k(g(h, n)^*\Gamma)$$

where  $\tilde{F}_k$  is compactly supported in the cylinder  $\pi^{-1}(B_k)$ .

If  $r$  is small enough, we have a smooth section  $\iota : B_k \rightarrow G$  that partially inverts the projection from  $G$  to  $G/(G_{(1,0)}\Gamma)$ , and so we can parameterise any element  $x$  of  $\pi^{-1}(B_k)$  uniquely as  $\iota(x_0)t\Gamma$  for some  $x_0 \in B_k$  and  $t \in T$  (noting that  $t\Gamma$  is well-defined as an element of  $G/\Gamma$ ). Similarly, we can parameterise any element of  ${}^*\pi^{-1}(B_k)$  uniquely as  $\iota(x_0)t\Gamma$  for  $x_0 \in {}^*B_k$  and  $t \in {}^*T$ .

We can now view the Lipschitz function  $F_k \in \text{Lip}^*(G/\Gamma)$  as a compactly supported Lipschitz function in  $\text{Lip}^*(B_k \times T)$ . Applying a Fourier (or Stone-Weierstrass) decomposition in the  $T$  directions (cf. Lemma E.5), we thus see that for any standard  $\varepsilon > 0$  we can approximate  $\tilde{F}_k$  uniformly to error  $\varepsilon/K$  by a sum  $\sum_{k'=1}^{K'} \tilde{F}_{k,k'}$ , where  $K'$  is standard and each  $F_{k,k'} \in \text{Lip}^*(B_k \times T)$  is compactly supported and has a character  $\xi_{k'} : T \rightarrow \mathbb{T}$  such that

$$\tilde{F}_{k,k'}(\iota(x_0)t\Gamma) = e(\xi_{k'}(t))\tilde{F}_{k,k'}(\iota(x_0)\Gamma) \quad (8.3)$$

for all  $x_0 \in {}^*(2B_k)$  and  $t \in {}^*T$ . It thus suffices to show that for each  $k, k'$ , the sequence

$$\tilde{\chi}_{k,k'} : (h, n) \mapsto \tilde{F}_{k,k'}(g(h, n)^*\Gamma)$$

is a linearised  $(1, s-1)$ -function.

Fix  $k, k'$ . Performing a Taylor expansion (Lemma B.9) of the polynomial sequence  $g \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}^2}^2 \rightarrow G_{\mathbb{N}^2})$ , we may write

$$g(h, n) = g_0(n)g_1(n)^h$$

where  $g_0 \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  is a one-dimensional polynomial map (giving  $G$  the  $\mathbb{N}$ -filtration  $G_{\mathbb{N}} := (G_{(i,0)})_{i \in \mathbb{N}}$ ), and  $g_1 \in {}^*\text{poly}(\mathbb{Z} \rightarrow (G_{(1,0)})_{\mathbb{N}})$  is another one-dimensional polynomial map (giving the abelian group  $G_{(1,0)}$  the  $\mathbb{N}$ -filtration  $(G_{(1,0)})_{\mathbb{N}} := (G_{(1,i)})_{i \in \mathbb{N}}$ ). In particular, we see that  $\tilde{\chi}_{k,k'}(h, n)$  is only non-vanishing when  $\pi(g_0(n)^*\Gamma) \in B$ . Furthermore, in that case we see from (8.3) that

$$\tilde{\chi}_{k,k'}(h, n) = e(h\xi(g_1(n) \bmod \Gamma_{(1,0)}))\tilde{F}_{k,k'}(g_0(n)^*\Gamma), \quad (8.4)$$

which gives the required factorisation (8.2) with  $c(n) := e(\xi(g_1(n) \bmod \Gamma_{(1,0)}))$  and  $\psi(n) := \tilde{F}_{k,k'}(g_0(n)^*\Gamma)$ .

The only remaining task is to establish that for any given  $h, l$ , the sequence  $n \mapsto c(n-l)^h \overline{c(n)}^h$  is a degree  $\leq s-2$  nilsequence. We expand this sequence as

$$n \mapsto e(h(\xi(g_1(n-l) \bmod \Gamma_{(1,0)}) - \xi(g_1(n) \bmod \Gamma_{(1,0)})))$$

But from the abelian nature of  $G_{(1,0)}$ , the map  $n \mapsto \xi(g_1(n) \bmod \Gamma_{(1,0)})$  is a polynomial map from  ${}^*\mathbb{Z}$  to  ${}^*\mathbb{T}$  of degree at most  $s-1$ , and the claim follows.  $\square$

We now return to the proof of Theorem 7.3. With this multiplicative structure, we can now begin the Cauchy-Schwarz argument. By hypothesis, for each  $h \in H$  we can find a scalar nilsequence  $\psi_h$  of degree  $\leq s-2$  such that

$$|\mathbb{E}_{n \in [N]} \Delta_h f(n) \overline{\chi(h, n)} \otimes \overline{\chi_h(n) \psi_h(n)}| \gg 1.$$

By Corollary A.12, we may ensure that  $\psi_h$  varies in a limit fashion on  $h$ . Applying Corollary A.6, this lower bound is uniform in  $h$ .

Applying Proposition 8.3 (with a sufficiently small  $\varepsilon$ ) and using the pigeonhole principle, we may then find a linearised  $(1, s-1)$ -function  $(h, n) \mapsto c(n)^h \psi(n)$  such that

$$|\mathbb{E}_{n \in [N]} \Delta_h f(n) c(n)^{-h} \overline{\psi(n)} \otimes \overline{\chi_h(n) \psi_h(n)}| \gg 1.$$

By Corollary A.6 again, the lower bound is still uniform in  $h$ . We may then average in  $h$  (extending  $\psi_h, \chi_h$  by zero for  $h$  outside of  $H$ ) and conclude that

$$\mathbb{E}_{h \in [N]} |\mathbb{E}_{n \in [N]} \Delta_h f(n) c(n)^{-h} \overline{\psi(n)} \otimes \overline{\chi_h(n) \psi_h(n)}| \gg 1,$$

thus there exists a scalar function  $b \in L^\infty[[N]]$  such that

$$|\mathbb{E}_{h \in [N]} \mathbb{E}_{n \in [N]} b(h) f(n+h) \overline{f(n)} c(n)^{-h} \overline{\psi(n)} \otimes \overline{\chi_h(n) \psi_h(n)}| \gg 1.$$

By absorbing  $b(h)$  into the  $\psi_h$  factor, we may now drop the  $b(h)$  factor. We write  $n+h=m$  and obtain

$$|\mathbb{E}_{m \in [N]} f(m) \mathbb{E}_{h \in [N]} c(m-h)^{-h} f'(m-h) \otimes \overline{\chi_h(m-h) \psi_h(m-h)}| \gg 1$$

where  $f' := \overline{f\psi}$  (recall that  $f$  is extended by zero outside of  $[N]$ ), which by Cauchy-Schwarz implies that

$$\begin{aligned} & |\mathbb{E}_{m \in [N]} \mathbb{E}_{h, h' \in [N]} c(m-h)^{-h} c(m-h')^{h'} f'(m-h) \otimes \overline{f(m-h')} \\ & \quad \otimes \overline{\chi_h(m-h)} \otimes \chi_{h'}(m-h') \overline{\psi_h(m-h)} \overline{\psi_{h'}(m-h')}| \gg 1. \end{aligned}$$

Making the change of variables  $h' = h + l$ ,  $n = m - h$ , we obtain

$$\begin{aligned} & |\mathbb{E}_{h, l \in [2N]; n \in [N]} c(n)^{-h} c(n-l)^{h+l} f'(n) \otimes \overline{f'(n-l)} \\ & \quad \otimes \overline{\chi_h(n)} \otimes \chi_{h+l}(n-l) \overline{\psi_h(n)} \overline{\psi_{h+l}(n-l)}| \gg 1. \end{aligned}$$

We then simplify this as

$$|\mathbb{E}_{h, l \in [2N]; n \in [N]} c_2(l, n) \otimes \overline{\chi_h(n)} \otimes \chi_{h+l}(n-l) \psi_{h,l}(n)| \gg 1 \quad (8.5)$$

where

$$\begin{aligned} c_2(l, n) &:= c(n-l)^l f'(n) \otimes \overline{f'(n-l)} \\ \psi_{h,l}(n) &= c(n-l)^h c(n)^{-h} \overline{\psi_h(n)} \psi_{h+l}(n-l) \end{aligned}$$

Clearly  $c_2$  is bounded. As for  $\psi_{h,l}$ , we see from Definition 8.1 and Corollary E.2 that  $\psi_{h,l}$  is a nilsequence of degree  $\leq s-2$  for each  $h, l$ .

Returning to (8.5), we use the pigeonhole principle to conclude that for many  $k \in [[2N]]$ , we have

$$|\mathbb{E}_{h \in [2N]; n \in [N]} c_2(k, n) \otimes \overline{\chi_h(n)} \otimes \chi_{h+k}(n-k) \psi_{h,k}(n)| \gg 1.$$

Let  $k$  be such that the above estimate holds. Applying Cauchy-Schwarz in the  $n$  variable to eliminate the  $c_2(k, n)$  term, we have

$$|\mathbb{E}_{h, h' \in [2N]; n \in [N]} \overline{\chi_h(n)} \otimes \chi_{h+k}(n-k) \otimes \overline{\chi_{h'}(n)} \otimes \chi_{h'+k}(n-k) \psi_{h,k}(n)| \gg 1$$

and thus for many  $k, h, h' \in [[2N]]$ , we have

$$|\mathbb{E}_{n \in [N]} \overline{\chi_h(n)} \otimes \chi_{h+k}(n-k) \otimes \overline{\chi_{h'}(n)} \otimes \chi_{h'+k}(n-k) \psi_{h,k}(n)| \gg 1,$$

which implies that

$$n \mapsto \overline{\chi_h(n)} \otimes \chi_{h+k}(n-k) \otimes \overline{\chi_{h'}(n)} \otimes \chi_{h'+k}(n-k)$$

is  $(s-2)$ -biased on  $[N]$ . Note that this forces  $h, h+k, h', h'+k$  to be an additive quadruple in  $H$ , as otherwise the expression vanishes. Applying a change of variables, we obtain Proposition 7.3.

For future reference we observe that a simpler version of the same argument (in which the  $\chi$  and  $\psi_h$  factors are not present) gives

**Proposition 8.4** (Cauchy-Schwarz). *Let  $f \in L^\infty[N]$ , let  $H$  be a dense subset of  $[N]$ , and suppose that one has a family of functions  $\chi_h \in L^\infty(*\mathbb{Z})$  depending in a limit fashion on  $h$ , such that  $\Delta_h f$  correlates with  $\chi_h$  on  $[N]$  for all  $h \in H$ . Then for many (i.e. for  $\gg N^3$ ) additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ , the sequence*

$$n \mapsto \chi_{h_1}(n) \otimes \chi_{h_2}(n + h_1 - h_4) \otimes \overline{\chi_{h_3}(n)} \otimes \overline{\chi_{h_4}(n + h_1 - h_4)} \quad (8.6)$$

*is biased.*

This proposition in fact has quite a simple proof; see [29]. Note how we can conclude (8.6) to be biased and not merely  $(s-2)$ -biased. As such, Proposition 8.4 saves some “lower order” information that was not present in Proposition 7.3; this lower order information will be crucial later in the argument, when we establish the symmetry property in Theorem 7.4.

## 9. FREQUENCIES AND REPRESENTATIONS

We will use Proposition 7.3 to analyse the “frequency” of the nilcharacters  $(\chi_h)_{h \in H}$  appearing in Theorem 7.2. To motivate the discussion, let us first suppose that we are in the (significantly simpler)  $s = 2$  case, rather than the actual case  $s \geq 3$  of interest. When  $s = 2$ , we can represent  $\chi_h$  as a linear phase  $\chi_h(n) = e(\xi_h n + \theta_h)$  for some  $\xi_h, \theta_h \in *\mathbb{T}$ ; one can then interpret  $\xi_h$  as the *frequency* of  $h$ .

In order to describe how this frequency  $\xi_h$  behaves in  $h$ , it will be convenient to *represent*  $\xi_h$  as a linear combination

$$\xi_h = a_{1,h}\xi_{1,h} + \dots + a_{D,h}\xi_{D,h} \quad (9.1)$$

of other frequencies  $\xi_{1,h}, \dots, \xi_{D,h} \in *\mathbb{T}$ , where the  $a_{i,h} \in \mathbb{Z}$  are (standard) integer coefficients, and the  $(\xi_{i,h})_{h \in H}$  are families of frequencies which have better properties with regards to their dependence on  $h$ ; for instance, they might be “core frequencies”  $\xi_{i,h} = \xi_{*,i}$  that are independent of  $h$ , or they might be “bracket-linear petal” frequencies that depend in a bracket-linear fashion on  $h$ , or they might be “regular petal” frequencies which behave in a suitably “dissociated” manner in  $h$ .

We can schematically depict the relationship (9.1) as

$$[\chi_h] \approx \eta_h(\mathcal{F}_h)$$

where  $[\chi_h]$  is some sort of “symbol” of  $\chi_h$  (which, in the linear case  $s = 2$ , is just  $\xi_h \bmod 1$ ),  $\mathcal{F}_h \in *\mathbb{T}^D$  is the *frequency vector*  $\mathcal{F}_h = (\xi_{1,h}, \dots, \xi_{D,h})$ , and  $\eta_h : *\mathbb{T}^D \rightarrow *\mathbb{T}$  is the *vertical frequency*

$$\eta_h(x_1, \dots, x_D) := a_{1,h}x_1 + \dots + a_{D,h}x_D. \quad (9.2)$$

We will need to find analogues of the above type of representation in higher degree  $s \geq 3$ . Heuristically, we will wish to represent the symbol  $[\chi]_{\Xi_{\text{DR}}^{(s-1, r_*)}([N])}$  of a nilcharacter  $\chi$  on  $[N]$  of degree-rank  $(s-1, r_*)$  (which will ultimately depend on a parameter  $h$ , though we will not need this parameter in the current discussion) heuristically as

$$[\chi]_{\Xi_{\text{DR}}^{(s-1, r_*)}([N])} \approx \eta(\mathcal{F}) \quad (9.3)$$

where  $\mathcal{F} = (\xi_{i,j})_{1 \leq i \leq s-1, 1 \leq j \leq D_i}$  is a *horizontal frequency vector* of frequencies  $\xi_{i,j} \in *\mathbb{T}$  associated to a *dimension vector*  $\vec{D} = (D_1, \dots, D_{s-1})$ , and  $\eta$  is a *vertical frequency* that generalises (9.2), but whose precise form we are not yet ready

to describe precisely. We then say that the triple  $(\vec{D}, \eta, \mathcal{F})$  forms a *total frequency representation* of  $\chi$ .

In the previous paper [28] that treated the  $s = 3$  case, such a representation was implicitly used via the description of degree-rank  $(2, 2)$  nilcharacters  $\chi_h$  as essentially being bracket quadratic phases  $e(\sum_{j=1}^J \{\alpha_{h,j}n\}\beta_{h,j}n)$  modulo lower order terms (and ignoring the issue of vector-valued smoothing for now). In our current language, this would correspond to a dimension vector  $\vec{D} = (2J, 0)$  and a horizontal frequency vector of the form  $(\alpha_{h,1}, \dots, \alpha_{h,J}, \beta_{h,1}, \dots, \beta_{h,J})$ , and a certain vertical frequency  $\eta$  depending only on  $J$  that we are not yet ready to describe explicitly here. Bracket-calculus identities such as (6.4) could then be used to manipulate such a universal frequency representation into a suitably “regularised” form.

In principle, one could also use bracket calculus to extract the symbol of  $\chi_h$  in terms of frequencies such as  $\alpha_{h,j}$  and  $\beta_{h,j}$  for higher values of  $s$ . However, as we are avoiding the use of bracket calculus machinery here, we will proceed instead using the language of nilmanifolds, and in particular by lifting the nilmanifold  $G_h/\Gamma_h$  up to a *universal nilmanifold* in order to obtain a suitable space (independent of  $h$ ) in which to detect relationships between frequencies such as  $\alpha_{h,j}, \beta_{h,j}$ . In some sense, this universal nilmanifold will play the role that the unit circle  $\mathbb{T}$  plays in Fourier analysis.

We first define the notion of universal nilmanifold that we need.

**Definition 9.1** (Universal nilmanifold). A *dimension vector* is a tuple

$$\vec{D} = (D_1, \dots, D_{s-1}) \in \mathbb{N}^{s-1}$$

of standard natural numbers. Given a dimension vector, we define the *universal nilpotent group*  $G^{\vec{D}} = G^{\vec{D}, \leq (s-1, r_*)}$  of degree-rank  $(s-1, r_*)$  to be the Lie group generated by formal generators  $e_{i,j}$  for  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_i$ , subject to the following constraints:

- Any  $(m-1)$ -fold iterated commutator of  $e_{i_1, j_1}, \dots, e_{i_m, j_m}$  with  $i_1 + \dots + i_m \geq s$  is trivial.
- Any  $(m-1)$ -fold iterated commutator of  $e_{i_1, j_1}, \dots, e_{i_m, j_m}$  with  $i_1 + \dots + i_m = s-1$  and  $m \geq r+1$  is trivial.

We give this group a degree-rank filtration  $(G_{(d,r)}^{\vec{D}})_{(d,r) \in \text{DR}}$  by defining  $G_{(d,r)}^{\vec{D}}$  to be the Lie group generated by  $(m-1)$ -fold iterated commutators of  $e_{i_1, j_1}, \dots, e_{i_m, j_m}$  with  $1 \leq i_l \leq s-1$  and  $1 \leq j_l \leq D_{i_l}$  for all  $1 \leq l \leq m$  for which either  $i_1 + \dots + i_m > d$ , or  $i_1 + \dots + i_m = d$  and  $m \geq r$ . It is not hard to verify that this is indeed a filtration of degree-rank  $\leq (s-1, r_*)$ . We then let  $\Gamma^{\vec{D}}$  be the discrete group generated by the  $e_{i,j}$  with  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_i$ , and refer to  $G^{\vec{D}}/\Gamma^{\vec{D}}$  as the *universal nilmanifold* with dimension vector  $\vec{D}$ .

A *universal vertical frequency* at dimension vector  $\vec{D}$  is a continuous homomorphism  $\eta : G_{(s-1, r_*)}^{\vec{D}} \rightarrow \mathbb{R}$  which sends  $\Gamma_{(s-1, r_*)}^{\vec{D}}$  to the integers (i.e. a filtered homomorphism from  $G_{(s-1, r_*)}^{\vec{D}}/\Gamma_{(s-1, r_*)}^{\vec{D}}$  to  $\mathbb{T}$ ).

*Remark.* One can give an explicit basis for this nilmanifold in terms of certain iterated commutators of the  $e_{i,j}$ , following [46, 49]. This can then be used to relate nilcharacters to bracket polynomials, as in [46], and it is then possible to develop enough of a “bracket calculus” to substitute for some of the nilpotent algebra performed in this paper. However, we will not proceed by such a route here



(as it would make the paper even longer than it currently is), and in fact will not need an explicit basis for universal nilmanifolds at all.

**Example 9.2.** The unit circle with the degree  $\leq d$  filtration (see Example 4.3) is isomorphic to the universal nilmanifold  $G^{(0,\dots,0,1),\leq(d,1)}$ , thus for instance the unit circle with the lower central series filtration is isomorphic to  $G^{(1),\leq(1,1)}$ . A universal vertical frequency for any of these nilmanifolds is essentially just a map of the form  $\eta : x \mapsto nx$  for some integer  $n$ .

**Example 9.3.** The Heisenberg group (6.1) (with the lower central series filtration) is the universal nilpotent group  $G^{(2,0)} = G^{(2,0),\leq(2,2)}$  of degree-rank  $(2,2)$  (after identifying  $e_1, e_2$  with  $e_{1,1}$  and  $e_{1,2}$  respectively), and the Heisenberg nilmanifold  $G/\Gamma$  is the corresponding universal nilmanifold  $G^{(2,0)}/\Gamma^{(2,0)}$ . If we reduce the degree-rank from  $(2,2)$  to  $(2,1)$ , then the commutator  $[e_1, e_2]$  now trivialises, and  $G^{(2,0),\leq(2,1)}$  collapses to the abelian Lie group  $\mathbb{R}^2 \equiv G^{2,\leq(1,1)}$ , with universal nilmanifold  $\mathbb{T}^2$ .

If, instead of the lower central series filtration, one gives the Heisenberg group (6.1) the filtration used in Example 6.21 to model the sequence (6.9), then this group is isomorphic to the universal nilpotent group  $G^{(1,1),\leq(3,2)}$ , with the two generators  $e_1, e_2$  of the Heisenberg group now being interpreted as  $e_{1,1}$  and  $e_{2,1}$  respectively.

**Example 9.4.** Consider the universal nilpotent group  $G^{(D_1,D_2,D_3),\leq(3,3)}$ . This group is generated by “degree 1” generators  $e_{1,1}, \dots, e_{1,D_1}$ , “degree 2” generators  $e_{2,1}, \dots, e_{2,D_2}$ , and “degree 3” generators  $e_{3,1}, \dots, e_{3,D_3}$ , with any iterated commutator of total degree exceeding three vanishing (thus for instance the degree 3 generators are central, and the degree 2 generators commute with each other). If one drops the degree-rank from  $(3,3)$  to  $(3,2)$ , then all triple commutators of degree 1-generators, such as  $[[e_{1,i}, e_{1,j}], e_{1,k}]$  now vanish, reducing the dimension of the nilpotent group. Dropping the degree-rank further to  $(3,1)$  also eliminates the commutators of degree 1 and degree 2 generators (thus making the degree 2 generators central). Finally, dropping the degree-rank to  $(3,0)$  eliminates the degree 3 generators completely, and indeed  $G^{(D_1,D_2,D_3),\leq(3,0)}$  is isomorphic to  $G^{(D_1,D_2),\leq(2,2)}$ .

**Example 9.5.** The free  $s$ -step nilpotent group on  $D$  generators, in our notation, becomes  $G^{(D,0,\dots,0),\leq(s,s)}$ . We may thus view the universal nilpotent groups  $G^{\vec{D},\leq(d,r)}$  as generalisations of the free nilpotent groups, in which some of the generators are allowed to be weighted to have degrees greater than 1, and there is an additional rank parameter to cut down some of the top-order behaviour.

It will be an easy matter to lift a nilcharacter  $\chi$  from a general degree-rank  $\leq (s-1, r_*)$  nilmanifold  $G/\Gamma$  to a universal nilmanifold  $G^{\vec{D}}/\Gamma^{\vec{D}}$  for some sufficiently large dimension vector  $\vec{D}$  (see Lemma 9.12 below). Once one does so, we will need to extract the various “top order frequencies” present in that nilcharacter. For instance, if  $s = 4$  and  $\chi$  is (some vector-valued smoothing of) the degree 3 phase

$$n \mapsto e(\{\alpha n\}\beta n^2 + \gamma n^3 + \delta n^2 + \{\epsilon n\}\mu n + \nu n + \theta)$$

then we will need to extract out the “degree 3” frequency  $\gamma$ , the “degree 2” frequency  $\beta$ , and the “degree 1” frequency  $\alpha$ . (The remaining parameters  $\delta, \epsilon, \mu, \nu, \theta$  only contribute to terms of degree strictly less than 3, and will not need to be extracted.)

As it turns out, the degree  $i$  frequencies will most naturally live in the  $i^{\text{th}}$  *horizontal torus* of the relevant universal nilmanifold; we now pause to define these torii precisely. (These torii also implicitly appeared in [27, Appendix A].)

**Definition 9.6** (Horizontal Taylor coefficients). Let  $G = (G, (G_{(d,r)})_{(d,r) \in \text{DR}})$  be a degree-rank-filtered nilpotent group. For every  $i \geq 0$ , define the  $i^{\text{th}}$  *horizontal space*  $\text{Horiz}_i(G)$  to be the abelian group

$$\text{Horiz}_i(G) := G_{(i,1)} / G_{(i,2)},$$

with the convention that  $G_{(d,r)} := G_{(d+1,0)}$  if  $r > d$  (so in particular,  $G_{(1,2)} = G_{(2,0)}$ ).

For any polynomial map  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ , we define the  $i^{\text{th}}$  *horizontal Taylor coefficient*  $\text{Taylor}_i(g) \in \text{Horiz}_i(G)$  to be the quantity

$$\text{Taylor}_i(g) := \partial_1 \dots \partial_1 g(n) \bmod G_{(i,2)}$$

for any  $n \in \mathbb{Z}$ . Note that this map is well-defined since  $\partial_1 \dots \partial_1 g$  takes values in  $G_{(i,1)}$  and has first derivatives in  $G_{(i+1,1)}$  and hence in  $G_{(i,2)}$ .

If  $\Gamma$  is a subgroup of  $G$ , we define

$$\text{Horiz}_i(G/\Gamma) := \text{Horiz}_i(G) / \text{Horiz}_i(\Gamma)$$

and for a polynomial orbit  $\mathcal{O} \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}}) := \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}}) / \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$ , we define the  $i^{\text{th}}$  *horizontal Taylor coefficient*  $\text{Taylor}_i(\mathcal{O}) \in \text{Horiz}_i(G/\Gamma)$  to be the quantity defined by

$$\text{Taylor}_i(g\Gamma) := \text{Taylor}_i(g) \bmod \text{Horiz}_i(\Gamma)$$

for any  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ ; it is easy to see that this quantity is well-defined.

These concepts extend to the ultralimit setting in the obvious manner; thus for instance, if  $\mathcal{O} \in {}^*\text{poly}(H_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$ , then  $\text{Taylor}_i(\mathcal{O})$  is an element to  ${}^*\text{Horiz}_i(G/\Gamma)$ .

If  $G/\Gamma$  is a degree-rank filtered nilmanifold, it is easy to see that the horizontal spaces  $\text{Horiz}_i(G)$  are abelian Lie groups, and that  $\text{Horiz}_i(\Gamma)$  is a sublattice of  $\text{Horiz}_i(G)$ , so  $\text{Horiz}_i(G/\Gamma)$  is a torus, which we call the  $i^{\text{th}}$  *horizontal torus* of  $G/\Gamma$ .

*Remark.* The above definition can be generalised by replacing the domain  $\mathbb{Z}$  with an arbitrary additive group  $H = (H, +)$ . In that case, the Taylor coefficient  $\text{Taylor}_i(g)$  is not a single element of  $\text{Horiz}_i(G)$ , but is instead a map  $\text{Taylor}_i(g) : H^i \rightarrow \text{Horiz}_i(G)$  defined by the formula

$$\text{Taylor}_i(g)(h_1, \dots, h_k) := \partial_{h_1} \dots \partial_{h_k} g(n) \bmod G_{(i,2)}$$

for  $h_1, \dots, h_k \in H$ . Using Corollary B.7 we easily see that this map is symmetric and multilinear; thus for instance when  $H = \mathbb{Z}$  we have

$$\text{Taylor}_i(g)(h_1, \dots, h_k) = h_1 \dots h_k \text{Taylor}_i(g).$$

However, we will not need this generalisation here.

A further application of Corollary B.7 shows that the map  $g \mapsto \text{Taylor}_i(g)$  is a homomorphism. As a corollary, we see that any translate  $g(\cdot + h) = (\partial_h g)g$  of  $g$  will have the same Taylor coefficients as  $g$ :  $\text{Taylor}_i(g(\cdot + h)) = \text{Taylor}_i(g)$ .

**Example 9.7.** Consider the unit circle  $G/\Gamma = \mathbb{T}$  with the degree  $\leq d$  filtration (see Example 4.3). Then the  $d^{\text{th}}$  horizontal torus is  $\mathbb{T}$ , and all other horizontal tori are trivial. If  $\alpha_0, \dots, \alpha_d \in {}^*\mathbb{R}$ , then the map  $\mathcal{O} : n \mapsto \alpha_0 + \dots + \alpha_d n^d \bmod 1$  is a polynomial orbit in  ${}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \mathbb{T}_{\mathbb{N}})$ , and the  $d^{\text{th}}$  horizontal Taylor coefficient is the quantity  $d!\alpha_d \bmod 1$  from  ${}^*\mathbb{Z}^d$  to  ${}^*\mathbb{T}$ . (All other horizontal Taylor coefficients are of course trivial.) Thus we see that the horizontal coefficient captures most of the top order coefficient  $\alpha_d$ , but totally ignores all lower order terms.

**Example 9.8.** Let  $G = G^{(2,1)} = G^{(2,1), \leq (2,2)}$  be the universal nilpotent group of degree-rank  $(2, 2)$ . Thus  $G$  is generated by  $e_{1,1}, e_{1,2}, e_{2,1}$ , with relations

$$[[e_{1,1}, e_{1,2}], e_{1,i}] = [e_{1,i}, e_{2,1}] = 1 \quad \text{for } i = 1, 2.$$

and with the degree-rank filtration

$$\begin{aligned} G_{(0,0)} &= G_{(1,0)} = G_{(1,1)} = G \\ G_{(2,0)} &= G_{(2,1)} = \langle [e_{1,1}, e_{1,2}], e_{2,1} \rangle_{\mathbb{R}} \\ G_{(2,2)} &= \langle [e_{1,1}, e_{1,2}] \rangle_{\mathbb{R}} \end{aligned}$$

and the lattice

$$\Gamma = \Gamma^{(2,2)} = \Gamma^{(2,2), \leq (2,1)} := \langle e_{1,1}, e_{1,2}, e_{2,1} \rangle.$$

Let  $\alpha, \beta, \gamma \in {}^*\mathbb{R}$ , and consider the orbit  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  defined by the formula

$$\mathcal{O}(n) := e_{1,1}^{n\alpha} e_{1,2}^{n\beta} e_{2,1}^{n^2\gamma};$$

this is polynomial by Example 4.4. Then

$$\text{Taylor}_1(g) = \partial_1 g(n) \bmod {}^*G_{(2,0)} = e_{1,1}^{\alpha} e_{1,2}^{\beta} \bmod {}^*G_{(2,0)},$$

and

$$\text{Taylor}_2(g) = e_{2,1}^{2\gamma} \bmod {}^*G_{(2,2)}.$$

Then  $\text{Taylor}_0(g(n){}^*\Gamma) = g(n){}^*\Gamma$ ,

$$\text{Taylor}_1(g{}^*\Gamma) = e_{1,1}^{\alpha} e_{1,2}^{\beta} \bmod G_{(2,0)} {}^*\Gamma$$

and

$$\text{Taylor}_2(g{}^*\Gamma) = e_{2,1}^{2\gamma} \bmod {}^*G_{(2,2)} \Gamma_{(2,0)}.$$

**Example 9.9.** Let  $G/\Gamma$  be the Heisenberg nilmanifold (6.1) with the lower central series filtration. Thus  $G/\Gamma$  is a degree  $\leq 2$  nilmanifold, which can then be viewed as a degree-rank  $\leq (2, 2)$  nilmanifold by Example 6.11. The first horizontal torus  $\text{Horiz}_1(G/\Gamma)$  is isomorphic to the 2-torus  $\mathbb{T}^2$ , with generators given by  $e_1, e_2 \bmod G_2\Gamma$ . The second horizontal torus  $\text{Horiz}_2(G/\Gamma)$  is trivial, since  $G_{(2,1)} = [G, G]$  is equal to  $G_{(2,0)} = G_2$ . If  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  is the polynomial orbit  $\mathcal{O} : n \mapsto e_2^{\beta n} e_1^{\alpha n} {}^*\Gamma$ , then the first Taylor coefficient is the quantity  $(\alpha, \beta)$ . Note also that if one modified the polynomial orbit by a further factor of  $[e_1, e_2]^{\gamma n^2 + \delta n + \epsilon}$ , this would not impact the Taylor coefficients at all. Thus we see that the Taylor coefficients only capture the frequencies associated to raw generators such as  $e_1$  and  $e_2$ , and not to commutators such as  $[e_1, e_2]$ .

**Example 9.10.** Now consider the Heisenberg group (6.1) with the filtration used in Example 6.21 to model the sequence (6.9). This is now a degree  $\leq 3$  nilmanifold, whose first horizontal torus  $\text{Horiz}_1(G/\Gamma)$  is isomorphic to the one-torus  $\mathbb{T}$  with generator  $e_2 \bmod G_{(2,0)}\Gamma$ , whose second horizontal torus  $\text{Horiz}_2(G/\Gamma)$  is isomorphic

to the one-torus  $\mathbb{T}$  with generator  $e_1 \bmod G_{(2,2)}\Gamma_{(2,1)}$ , and whose third horizontal torus  $\text{Horiz}_3(G/\Gamma)$  is trivial. If  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  is the polynomial orbit  $\mathcal{O} : n \mapsto e_2^{\beta n} e_1^{\alpha n^2} \Gamma$ , then the first Taylor coefficient is the linear limit map  $n \mapsto \beta n \bmod 1$ , and the second Taylor coefficient is the quantity  $2!\alpha \bmod 1$ .

We now have enough notation to be able to formally assign frequencies to a nilcharacter, by means of a package of data which we shall call a *representation*.

**Definition 9.11** (Representation). Let  $\chi \in L^\infty[N]$  be a nilcharacter of degree-rank  $\leq (s-1, r_*)$ . A *representation* of  $\chi$  is a collection of the following data:

- (i) A filtered nilmanifold  $G/\Gamma$  of degree-rank  $\leq (s-1, r_*)$ ;
- (ii) A filtered nilmanifold  $G_0/\Gamma_0$  of degree-rank  $\leq (s-1, r_*-1)$ ;
- (iii) A function  $F \in \text{Lip}^*(G/\Gamma \times G_0/\Gamma_0) \rightarrow \overline{S^\omega}$ ;
- (iv) Polynomial orbits  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  and  $\mathcal{O}_0 \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G_0/\Gamma_0)_{\mathbb{N}})$ ;
- (v) A dimension vector  $\vec{D} = (D_1, \dots, D_{s-1}) \in \mathbb{N}^{s-1}$ ;
- (vi) A universal vertical frequency  $\eta : G_{(s-1, r_*)}^{\vec{D}} \rightarrow \mathbb{R}$  at dimension  $\vec{D}$  on the universal nilmanifold  $G^{\vec{D}}/\Gamma^{\vec{D}}$  of degree-rank  $(s-1, r_*)$ ;
- (vii) A filtered homomorphism  $\phi : G^{\vec{D}}/\Gamma^{\vec{D}} \rightarrow G/\Gamma$  (see Definition 6.15);
- (viii) A *horizontal frequency vector*  $\mathcal{F} = (\xi_{i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_i}$  of frequencies  $\xi_{i,j} \in {}^*\mathbb{T}$ .

which obeys the following properties:

- (i) For all  $n \in [N]$ , one has

$$\chi(n) = F(\mathcal{O}(n), \mathcal{O}_0(n)). \quad (9.4)$$

- (ii) For every  $t \in G_{(s-1, r_*)}^{\vec{D}}$ , all  $x \in G/\Gamma$ , and  $x_0 \in G_0/\Gamma_0$ , one has

$$F(\phi(t)x, x_0) = e(\eta(t))F(x, x_0). \quad (9.5)$$

- (iii) For every  $1 \leq i \leq s-1$ , one has

$$\text{Taylor}_i(\mathcal{O}) = \pi_{\text{Horiz}_i(G/\Gamma)} \left( \phi \left( \prod_{j=1}^{D_i} e_{i,j}^{\xi_{i,j}} \right) \right), \quad (9.6)$$

where  $\pi_{\text{Horiz}_i(G/\Gamma)} : G_i \rightarrow \text{Horiz}_i(G/\Gamma)$  is the projection map; observe that the right-hand side is well-defined even though  $\xi_{i,j}$  is only defined modulo 1.

We call the triplet  $(\vec{D}, \mathcal{F}, \eta)$  a *total frequency representation* of the nilcharacter  $\chi$ .

This is a rather complicated definition, and we now illustrate it with a number of examples. We begin with the  $s=2, r_*=1$  case, taking  $\chi$  to be the degree-rank  $(1, 1)$  nilcharacter

$$\chi(n) := e(\xi n + \theta)$$

for some  $\xi, \theta \in {}^*\mathbb{R}$ . Let  $D_1 \geq 1$  be an integer, let  $\mathcal{F} = (\xi_{1,1}, \dots, \xi_{1,D_1}) \in {}^*\mathbb{T}^{D_1}$  be a collection of frequencies, and let  $\eta : \mathbb{R}^{D_1} \rightarrow \mathbb{R}$  be the universal vertical frequency  $\eta(x_1, \dots, x_{D_1}) := a_1 x_1 + \dots + a_{D_1} x_{D_1}$  for some integers  $a_1, \dots, a_{D_1} \in \mathbb{Z}$ . Then  $((D_1), \mathcal{F}, \eta)$  will be a total frequency representation of  $\chi$  if  $\xi = a_1 \xi_{1,1} + \dots + a_{D_1} \xi_{1,D_1}$ . Indeed, in that case, one can take  $G/\Gamma = \mathbb{T}$  (with the degree-rank  $\leq (1, 1)$  filtration,

see Example 6.12),  $G_0/\Gamma_0$  to be trivial,  $F$  equal to the exponential function  $(x, ()) \mapsto e(x)$ ,  $\phi : \mathbb{T}^{D_1} \rightarrow \mathbb{T}$  to be the filtered homomorphism

$$\phi(x_1, \dots, x_{D_1}) := a_1 x_1 + \dots + a_{D_1} x_{D_1},$$

and  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \mathbb{T}_{\mathbb{N}})$  to be the orbit  $n \mapsto \xi n + \theta \bmod 1$ . This should be compared with (9.3) and the discussion at the start of the section.

For a slightly more complicated example, we take  $s = 3$ ,  $r_* = 1$ , and let  $\chi$  be the degree-rank  $(2, 1)$  nilcharacter

$$\chi(n) := e(\alpha n^2 + \beta n + \gamma).$$

We let  $D_2 \geq 1$  be an integer, set  $D_1 := 0$ , let  $\mathcal{F} = (( ), (\xi_{2,1}, \dots, \xi_{2,D_2})) \in {}^*\mathbb{T}^0 \times {}^*\mathbb{T}^{D_2}$  be a collection of frequencies, and let  $\eta : \mathbb{R}^{D_2} \rightarrow \mathbb{R}$  be the universal vertical frequency  $\eta(x_1, \dots, x_{D_2}) := a_1 x_1 + \dots + a_{D_2} x_{D_2}$  for some integers  $a_1, \dots, a_{D_2} \in \mathbb{Z}$ . Then  $((0, D_2), \mathcal{F}, \eta)$  will be a total frequency representation of  $\chi$  if  $\xi = a_1 \xi_{2,1} + \dots + a_{D_2} \xi_{2,D_2}$  (cf. (9.3)). Indeed, we can take  $G/\Gamma = \mathbb{T}$  with the degree-rank  $\leq (2, 1)$  filtration (see Example 6.12),  $G_0/\Gamma_0 = \mathbb{T}$  with the degree-rank  $\leq (1, 1)$  filtration, the orbit

$$\mathcal{O}(n) := (\alpha n^2 \bmod 1, \beta n + \gamma \bmod 1)$$

and  $F : G/\Gamma \times G_0/\Gamma_0 \rightarrow S^1$  to be the function

$$F(x, y) := e(x)e(y),$$

and  $\phi : \mathbb{T}^{D_2} \rightarrow \mathbb{T}$  to be the filtered homomorphism

$$\phi(x_1, \dots, x_{D_2}) := a_1 x_1 + \dots + a_{D_2} x_{D_2}.$$

Note how the lower order terms  $\beta n + \gamma$  in the phase of  $\chi$  are shunted off to the lower degree-rank nilmanifold  $G_0/\Gamma_0$  and thus do not interact at all with the data  $\mathcal{F}, \eta$ . In this particular case, this shunting off was unnecessary, and one could have easily folded these lower order terms into the dynamics of the primary nilmanifold  $G/\Gamma$ ; but in the next example we give, the lower order behaviour does genuinely need to be separated from the top order behaviour by placing it in a separate nilmanifold.

We now turn to a genuinely non-abelian example of a universal representation. For this, we take  $s = 3$ ,  $r_* = 2$ , and let  $\chi$  be a degree-rank  $(2, 2)$  nilcharacter that is a suitable vector-valued smoothing of the bracket polynomial phase

$$n \mapsto e(\{\alpha n\}\beta n + \gamma n^2).$$

We can express this nilcharacter as

$$\chi(n) = F(\mathcal{O}(n), \mathcal{O}_0(n)),$$

where  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  is the orbit

$$\mathcal{O}(n) := e_2^{\beta n} e_1^{\alpha n} \Gamma$$

into the Heisenberg nilmanifold (6.1) (which we give the degree-rank  $\leq (2, 2)$  filtration),  $\mathcal{O}_0 \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  is the orbit

$$\mathcal{O}_0(n) := \gamma n^2 \bmod 1$$

into the unit circle  $G_0/\Gamma_0 = \mathbb{T}$  (which we give the degree-rank  $\leq (2, 1)$  filtration, see Example 6.12), and  $F$  is a suitable vector-valued smoothing of the map

$$(e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}} \Gamma, y) \mapsto e(t_{12})e(y)$$

for  $t_1, t_2, t_{12} \in I_0$ . By Example 9.9, we have  $\text{Taylor}_1(\mathcal{O}) = (\alpha \bmod 1, \beta \bmod 1)$  and  $\text{Taylor}_2(\mathcal{O})$  is trivial.

Now let  $D_1 \geq 1$  be an integer, set  $D_2 := 0$ , let  $\mathcal{F} = ((\xi_{1,1}, \dots, \xi_{1,D_1}), ()) \in {}^*\mathbb{T}^{D_1} \times {}^*\mathbb{T}^0$  be a collection of frequencies. The subgroup  $G_{(2,2)}^{(D_1,0)}$  of the universal nilmanifold  $G^{(D_1,0)} = G^{(D_1,0), \leq (2,2)}$  is then the abelian Lie group generated by the commutators  $[e_{1,i}, e_{1,j}]$  for  $1 \leq i < j \leq D_1$ . We let  $a_1, \dots, a_{D_1}, b_1, \dots, b_{D_1} \in \mathbb{Z}$  be integers, and let  $\phi : G^{(D_1,0)}/\Gamma^{(D_1,0)} \rightarrow G/\Gamma$  be the filtered homomorphism that maps  $e_{1,i}$  to  $e_1^{a_i} e_2^{b_i}$  for  $i = 1, \dots, D_1$ , thus

$$\begin{aligned} & \phi\left(\prod_{i=1}^{D_1} e_{1,i}^{t_i} \prod_{1 \leq i < j \leq D_1} [e_{1,i}, e_{1,j}]^{t_{i,j}} \Gamma^{(D_1,0)}\right) \\ &= \prod_{i=1}^{D_1} (e_1^{a_i} e_2^{b_i})^{t_i} \prod_{1 \leq i < j \leq D_1} [e_1^{a_i} e_2^{b_i}, e_1^{a_j} e_2^{b_j}]^{t_{i,j}} \Gamma \\ &= e_1^{\sum_{i=1}^{D_1} a_i t_i} e_2^{\sum_{i=1}^{D_1} b_i t_i} [e_1, e_2]^{-\sum_{i=1}^{D_1} a_i b_i \binom{t_i}{2} - \sum_{1 \leq i < j \leq D_1} b_i a_j t_i t_j + \sum_{1 \leq i < j \leq D_1} (a_i b_j - a_j b_i) t_{i,j}} \Gamma. \end{aligned}$$

Let us now see what conditions are required for  $((D_1, 0), \eta, \mathcal{F})$  to be a total frequency representation of  $\chi$ . The condition (9.6) becomes the constraints

$$\begin{aligned} \alpha &= \sum_{i=1}^{D_1} a_i \xi_{1,i} \\ \beta &= \sum_{i=1}^{D_1} b_i \xi_{1,i}, \end{aligned}$$

while the condition (9.5) becomes

$$\eta([e_{1,i}, e_{1,j}]) = a_i b_j - a_j b_i \quad (9.7)$$

for all  $1 \leq i < j \leq D_1$ , or equivalently

$$\eta\left(\prod_{1 \leq i < j \leq D_1} [e_{1,i}, e_{1,j}]^{t_{i,j}}\right) = \sum_{1 \leq i < j \leq D_1} (a_i b_j - a_j b_i) t_{i,j}$$

Conversely, with these constraints we obtain a total frequency representation of  $\chi$  by  $((D_1, 0), \eta, \mathcal{F})$ . This should be compared with the heuristic (9.3). (Note from (6.4) that the top order component  $\{\alpha n\} \beta n$  of  $\chi$  is morally anti-symmetric in  $\alpha, \beta$  modulo lower order terms, which is consistent with the anti-symmetry observed in (9.7).) Note also that the term  $\gamma n^2$ , which has lesser degree-rank than the top order term  $\{\alpha n\} \beta n$ , plays no role, due to it being shunted off to the lower degree-rank nilmanifold  $G_0/\Gamma_0$ . If instead we placed this term as part of the principal nilmanifold, then this would create a non-trivial second Taylor coefficient  $\text{Taylor}_2(\mathcal{O})$  which would then require a non-zero value of  $D_2$  in order to recover a total frequency representation. Thus we see that in order to neglect terms of lesser degree-rank (but equal degree) it is necessary to create the secondary nilmanifold  $G_0/\Gamma_0$  as a sort of “junk nilmanifold” to hold all such terms.

We make the easy remark that every nilcharacter  $\chi$  of degree-rank  $\leq (s-1, r_*)$  has at least one representation.

**Lemma 9.12** (Existence of representation). *Let  $\chi$  be a nilcharacter of degree-rank  $(s-1, r_*)$  on  $[N]$ . Then there exists at least one total frequency representation  $(\vec{D}, \mathcal{F}, \eta)$  of  $\chi$ .*

*Proof.* By definition,  $\chi = F \circ \mathcal{O}$  for some degree-rank  $\leq (s-1, r_*)$  nilmanifold  $G/\Gamma$ , some  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$ , and some  $F \in \text{Lip}^*(G/\Gamma)$  with a vertical frequency. For each  $1 \leq i \leq s-1$ , let  $f_{i,1}, \dots, f_{i,D_i}$  be a basis of generators for  $\Gamma_i$ , and let  $\vec{D} := (D_1, \dots, D_{s-1})$  be the associated dimension vector. Then we have a filtered homomorphism  $\phi : G^{\vec{D}} \rightarrow G$  which maps  $e_{i,j}$  to  $f_{i,j}$  for all  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_i$ . It is easy to see that  $\phi$  is surjective from  $G_i^{\vec{D}}$  to  $G_i$  for each  $i$ , and so the map  $\pi_{\text{Horiz}_i(G/\Gamma)} \circ \phi$  is surjective from  $G_i^{\vec{D}}$  to  $\text{Horiz}_i(G/\Gamma)$ . It is now an easy matter to locate frequencies  $\xi_{i,j}$  obeying (9.6), and the vertical frequency property of  $F$  can be pulled back via  $\phi$  to give (9.5). Setting  $G_0/\Gamma_0$  to be trivial, we obtain the claim.  $\square$

To conclude this section, we now give some basic facts about total frequency representations. These facts will not actually be used in this paper, but may serve to consolidate one's intuition about the nature of these representations. We first observe some linearity in the vertical frequency  $\eta$ .

**Lemma 9.13** (Linearity). *Suppose that  $\chi, \chi'$  are two nilcharacters of degree-rank  $(s-1, r_*)$  on  $[N]$  that have total frequency representations  $(\vec{D}, \mathcal{F}, \eta)$  and  $(\vec{D}, \mathcal{F}, \eta')$  respectively. Then  $\overline{\chi}$  has a total frequency representation  $(\vec{D}, \mathcal{F}, -\eta)$ , and  $\chi \otimes \chi'$  has a total frequency representation  $(\vec{D}, \mathcal{F}, \eta + \eta')$ .*

*Proof.* This is a routine matter of chasing down the definitions, and noting that nilmanifolds, polynomial orbits, etc. behave well with respect to direct sums.  $\square$

**Lemma 9.14** (Triviality). *Suppose that  $\chi$  is a nilcharacter of degree-rank  $(s-1, r_*)$  on  $[N]$  that has a total frequency representation  $(\vec{D}, \mathcal{F}, 0)$ . Then  $\chi$  is a nilsequence of degree-rank  $\leq (s-1, r_* - 1)$  (i.e.  $[\chi]_{\text{Symb}_{\text{DR}}^{(s-1, r_*)}([N])} = 0$ ).*

*Proof.* By construction, we have

$$\chi(n) = F(\mathcal{O}(n), \mathcal{O}_0(n))$$

for some limit polynomial orbits  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$ ,  $\mathcal{O}_0 \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G_0/\Gamma_0)_{\mathbb{N}})$  into filtered nilmanifolds  $G/\Gamma, G_0/\Gamma_0$  of degree-rank  $\leq (s-1, r_*)$  and  $\leq (s-1, r_* - 1)$  respectively, where  $F \in \text{Lip}^*(G/\Gamma \times G_0/\Gamma_0 \rightarrow \overline{S}^{\omega})$ . Furthermore, there exists a filtered homomorphism  $\phi : G^{\vec{D}}/\Gamma^{\vec{D}} \rightarrow G/\Gamma$  such that (9.6) holds, and such that

$$F(\phi(t)x, x_0) = F(x, x_0). \quad (9.8)$$

for all  $t \in G_{(s-1, r_*)}^{\vec{D}}$ .

Let  $T$  be the closure of the set  $\{\phi(t) \bmod \Gamma_{(s-1, r_*)} : t \in G_{(s-1, r_*)}^{\vec{D}}\}$ ; this is a subtorus of the torus  $G_{(s-1, r_*)}/\Gamma_{(s-1, r_*)}$ , and thus acts on  $G/\Gamma$ . As  $F$  is continuous and obeys the invariance (9.8), we see that  $F$  is  $T$ -invariant; we may thus quotient out by  $T$  and assume that  $T$  is trivial. In particular,  $\phi$  now annihilates  $G_{(s-1, r_*)}^{\vec{D}}$ .

We give  $G$  a new degree-rank filtration  $(G'_{(d,r)})_{(d,r) \in \text{DR}}$  (smaller than the existing filtration  $(G_{(d,r)})_{(d,r) \in \text{DR}}$ ), by defining  $G'_{(d,r)}$  to be the connected subgroup of  $G$  generated by  $G_{(d,r+1)}$  (recalling the convention  $G_{(d,r)} := G_{(d+1,0)}$  when  $r > d$ ) together with the image  $\phi(G_{(d,r)}^{\vec{D}})$  of  $G_{(d,r)}^{\vec{D}}$ . It is easy to see that this is still a filtration, and that  $G/\Gamma$  remains a filtered nilmanifold with this filtration, but now the degree-rank is  $\leq (s-1, r_* - 1)$  rather than  $\leq (s-1, r_*)$ . Furthermore, from

(9.6) we see that  $\mathcal{O}$  is still a polynomial orbit with respect to this new filtration. As such,  $\chi$  is a nilsequence of degree-rank  $\leq (s-1, r_*-1)$  as required.  $\square$

Combining the above two lemmas we obtain the following corollary.

**Corollary 9.15** (Representation determines symbol). *Suppose that  $\chi, \chi'$  are two nilcharacters of degree-rank  $(s-1, r_*)$  on  $[N]$  that have a common total frequency representation  $(\vec{D}, \mathcal{F}, \eta)$ . Then  $\chi, \chi'$  are equivalent. In other words, the symbol  $[\chi]_{\Xi^{(s-1, r_*)}([N])}$  depends only on  $(\vec{D}, \mathcal{F}, \eta)$ .*

Note that the above results are consistent with the heuristic (9.3).

## 10. LINEAR INDEPENDENCE AND THE SUNFLOWER LEMMA

A basic fact of linear algebra is that every finitely generated vector space is finite-dimensional. In particular, if  $v_1, \dots, v_l$  are a finite collection of vectors in a vector space  $V$  over a field  $k$ , then there exists a finite linearly independent set of vectors  $v'_1, \dots, v'_{l'}$  in  $V$  such that each of the vectors  $v_1, \dots, v_l$  is a linear combination (over  $k$ ) of the  $v'_1, \dots, v'_{l'}$ . Indeed, one can take  $v'_1, \dots, v'_{l'}$  to be a set of vectors generating  $v_1, \dots, v_l$  for which  $l'$  is minimal, since any linear relation amongst the  $v'_1, \dots, v'_{l'}$  can be used to decrease<sup>5</sup> the “rank”  $l'$ , contradicting minimality (cf. the proof of classical Steinitz exchange lemma in linear algebra).

We will need analogues of this type of fact for frequencies  $\xi_1, \dots, \xi_l$  in the limit unit circle  ${}^*\mathbb{T}$ . However, this space is not a vector space over a field, but is merely a module over a commutative ring  $\mathbb{Z}$ . As such, the direct analogue of the above statement fails; indeed, any standard rational in  ${}^*\mathbb{T}$ , such as  $\frac{1}{2} \bmod 1$ , clearly cannot be represented as a linear combination (over  $\mathbb{Z}$ ) of a finite collection of frequencies in  ${}^*\mathbb{T}$  that are linearly independent over  $\mathbb{Z}$ .

However, the standard rationals are the *only* obstruction to the above statement being true. More precisely, we have

**Lemma 10.1** (Baby regularity lemma). *Let  $l \in \mathbb{N}$ , and let  $\xi_1, \dots, \xi_l \in {}^*\mathbb{T}$ . Then there exists  $l', l'' \in \mathbb{N}$  and  $\xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''} \in {}^*\mathbb{T}$  such that  $\xi'_1, \dots, \xi'_{l'}$  are linearly independent over  $\mathbb{Z}$  (i.e. there exist no standard integers  $a_1, \dots, a_{l'}$ , not all zero, such that  $a_1 \xi'_1 + \dots + a_{l'} \xi'_{l'} = 0$ ), each of the  $\xi''_i$  are rational (i.e. they live in  $\mathbb{Q} \bmod 1$ ), and each of the  $\xi_1, \dots, \xi_l$  are linear combinations (over  $\mathbb{Z}$ ) of the  $\xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}$ .*

*Proof.* Fix  $l, \xi_1, \dots, \xi_l$ . Define a *partial solution* to be a collection of objects  $l', l'', \xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}$  satisfying all of the required properties, except possibly for the linear independence of the  $\xi'_1, \dots, \xi'_{l'}$ . Clearly at least one partial solution exists, since one can take  $l' := l, l'' := 0$ , and  $\xi'_i := \xi_i$  for all  $1 \leq i \leq l$ . Now let  $l', l'', \xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}$  be a partial solution for which  $l'$  is minimal. We claim that  $\xi'_1, \dots, \xi'_{l'}$  is linearly independent over  $\mathbb{Z}$ , which will give the lemma. To see this, suppose for contradiction that there existed  $a_1, \dots, a_{l'} \in \mathbb{Z}$ , not all zero, such

<sup>5</sup>Indeed, one can recast this argument as a rank reduction argument instead of a minimal rank argument, for the same reason that the principle of infinite descent is logically equivalent to the well-ordering principle. In this infinitary (ultralimit) setting, there is very little distinction between the two approaches, although the minimality approach allows for slightly more compact notation and proofs. But in the finitary setting, it becomes significantly more difficult to implement the minimality approach, and the rank reduction approach becomes preferable. See [28] for finitary “rank reduction” style arguments analogous to those given here.



that  $a_1\xi'_1 + \dots + a_{l'}\xi'_{l'} = 0$ . Without loss of generality we may assume that  $a_1$  is non-zero. For each  $2 \leq j \leq l'$ , let  $\tilde{\xi}'_j \in {}^*\mathbb{T}$  be such that  $a_1\tilde{\xi}'_j = \xi'_j$ . We then have

$$\xi'_1 = - \sum_{j=2}^{l'} \frac{a_j}{a_1} \xi'_j + q \pmod{1}$$

for some standard rational  $q \in \mathbb{Q}$ . If we then replace  $\xi'_1, \dots, \xi'_{l'}$  by  $\tilde{\xi}'_2, \dots, \tilde{\xi}'_{l'}$  (decrementing  $l'$  to  $l' - 1$ ) and append  $q$  to  $\xi''_1, \dots, \xi''_{l''}$ , then we obtain a new partial solution with a smaller value of  $l'$ , contradicting minimality. The claim follows.  $\square$

This lemma is too simplistic for our applications, and we will need to modify it in a number of ways. The first is to introduce an error term.

**Definition 10.2** (Linear independence). Let  $\varepsilon > 0$  be a limit real, and let  $l \in \mathbb{N}$ . A set of frequencies  $\xi_1, \dots, \xi_l \in {}^*\mathbb{T}$  is said to be *independent modulo  $O(\varepsilon)$*  if there do not exist any collection  $a_1, \dots, a_l \in \mathbb{Z}$  of standard integers, not all zero, for which

$$a_1\xi_1 + \dots + a_l\xi_l = O(\varepsilon) \pmod{1}$$

(Thus, for instance, the empty set (with  $k = 0$ ) is trivially independent modulo  $O(\varepsilon)$ .) Equivalently,  $\xi_1, \dots, \xi_l$  are linearly independent over  $\mathbb{Z}$  after quotienting out by the subgroup  $\varepsilon\overline{\mathbb{R}} \pmod{1}$ .

This definition is only non-trivial when  $\varepsilon$  is an infinitesimal (i.e.  $\varepsilon = o(1)$ ). In practice,  $\varepsilon$  will be a negative power of the unbounded integer  $N$ .

We have the following variant of Lemma 10.1.

**Lemma 10.3** (Regularising one collection of frequencies). *Let  $l \in \mathbb{N}$ , let  $\xi_1, \dots, \xi_l \in {}^*\mathbb{T}$ , and let  $\varepsilon > 0$  be a limit real. Then there exist  $l', l'', l''' \in \mathbb{N}$  and*

$$\xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}, \xi'''_1, \dots, \xi'''_{l'''} \in {}^*\mathbb{T}$$

*such that  $\xi'_1, \dots, \xi'_{l'}$  are linearly independent modulo  $O(\varepsilon)$ , each of the  $\xi''_i$  are rational, each of the  $\xi'''_i$  are  $O(\varepsilon)$ , and each of the  $\xi_1, \dots, \xi_l$  are linear combinations (over  $\mathbb{Z}$ ) of the  $\xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}, \xi'''_1, \dots, \xi'''_{l'''}$ .*

One can view Lemma 10.1 as the degenerate case  $\varepsilon = 0$  of the above lemma.

*Proof.* We repeat the proof of Lemma 10.1. Define a *partial solution* to be a collection of objects  $l', l'', l''', \xi'_1, \dots, \xi'_{l'}, \xi''_1, \dots, \xi''_{l''}, \xi'''_1, \dots, \xi'''_{l'''}$  obeying all the required properties except possibly for the linear independence property. Again it is clear that at least one partial solution exists, so we may find a partial solution for which  $l'$  is minimal. We claim that this is a complete solution. For if this is not the case, we have

$$a_1\xi'_1 + \dots + a_{l'}\xi'_{l'} = O(\varepsilon) \pmod{1}$$

for some  $a_1, \dots, a_{l'} \in \mathbb{Z}$ , not all zero. Again, we may assume that  $a_1 \neq 0$ . We again select  $\tilde{\xi}'_2, \dots, \tilde{\xi}'_{l'} \in {}^*\mathbb{T}$  with  $a_1\tilde{\xi}'_j = \xi'_j$  for all  $2 \leq j \leq l'$ , and observe that

$$\xi'_1 = - \sum_{j=2}^{l'} \frac{a_j}{a_1} \xi'_j + q + s \pmod{1}$$

for some standard rational  $q \in \mathbb{Q}$  and some  $s = O(\varepsilon)$ . If we then replace  $\xi'_1, \dots, \xi'_{l'}$  by  $\tilde{\xi}'_2, \dots, \tilde{\xi}'_{l'}$ , and append  $q$  and  $s$  to  $\xi''_1, \dots, \xi''_{l''}$  and  $\xi'''_1, \dots, \xi'''_{l'''}$  respectively, we contradict minimality, and the claim follows.  $\square$

This lemma is still far too simplistic for our needs, because we will not be needing to regularise just one collection  $\xi_1, \dots, \xi_l$  of frequencies, but a whole *family*  $\xi_{h,1}, \dots, \xi_{h,l}$  of frequencies, where  $h$  ranges over a parameter set  $H$ . Such frequencies can exhibit a range of behaviour in  $h$ ; at one extreme, they might be completely independent of  $h$ , while at the other extreme, the frequencies may vary substantially as  $h$  does. It turns out that in some sense, the general case is a combination of these extreme cases.

In this direction we have the following stronger version of Lemma 10.3.

**Lemma 10.4** (Regularising many collections of frequencies). *Let  $l \in \mathbb{N}$ , let  $\varepsilon > 0$  be a limit real, let  $H$  be a limit finite set, and for each  $h \in H$ , let  $\xi_{h,1}, \dots, \xi_{h,l}$  be frequencies in  ${}^*\mathbb{T}$  that depend in a limit fashion on  $h$ . Then there exists a dense subset  $H'$  of  $H$ , standard natural numbers,  $l_*, l', l'', l''' \in \mathbb{N}$ , “core” frequencies  $\xi_{*,1}, \dots, \xi_{*,l_*}, \xi''_{*,1}, \dots, \xi''_{l'',*} \in {}^*\mathbb{T}$ , and “petal” frequencies*

$$\xi'_{h,1}, \dots, \xi'_{h,l'}, \xi'''_{h,1}, \dots, \xi'''_{h,l'''} \in {}^*\mathbb{T}$$

*for each  $h \in H'$  depending in a limit fashion on  $h$ , and obeying the following properties:*

- (i) (Independence) *For almost all triples  $(h_1, h_2, h_3) \in (H')^3$  (i.e. for all but  $o(|H'|^3)$  such triples), the frequencies*

$$\xi_{*,1}, \dots, \xi_{*,l_*}, \xi'_{h_1,1}, \dots, \xi'_{h_1,l'}, \xi'_{h_2,1}, \dots, \xi'_{h_2,l'}, \xi'_{h_3,1}, \dots, \xi'_{h_3,l'}$$

*are linearly independent modulo  $O(\varepsilon)$ .*

- (ii) (Rationality) *For each  $1 \leq j \leq l''$ ,  $\xi''_{*,j}$  is a standard rational.*

- (iii) (Smallness) *For each  $h \in H'$  and  $1 \leq j \leq l'''$ ,  $\xi'''_{h,j} = O(\varepsilon)$ .*

- (iv) (Representation) *For each  $h \in H'$ , the  $\xi_{h,1}, \dots, \xi_{h,l}$  are linear combinations over  $\mathbb{Z}$  of the frequencies*

$$\xi_{*,1}, \dots, \xi_{*,l_*}, \xi'_{h,1}, \dots, \xi'_{h,l'}, \xi''_{*,1}, \dots, \xi''_{*,l''}, \xi'''_{h,1}, \dots, \xi'''_{h,l'''}$$

Note that Lemma 10.4 collapses to Lemma 10.3 if  $H$  is a singleton set.

*Proof.* We again use the usual argument. Define a *partial solution* to be a collection of objects  $H', l_*, l', l'', l''', \xi_{*,j}, \xi'_{h,j}, \xi''_{*,j}, \xi'''_{h,j}$  obeying all the required properties except possibly for the independence property. Again, at least one partial solution exists, since we may take  $H' := H$ ,  $l_* := l'' := l''' := 0$ ,  $l' := l$ , and  $\xi'_{j,h} := \xi_{j,h}$  for all  $h \in H$  and  $1 \leq j \leq l$ . We may thus select a partial solution for which  $l'$  is minimal; and among all such partial solutions with  $l'$  minimal, we choose a solution with  $l_*$  minimal for fixed  $l'$  (i.e. we minimise with respect to the lexicographical ordering on  $l'$  and  $l_*$ ). We claim that this doubly minimal solution obeys the independence property, which would give the claim.

Suppose the independence property fails. Carefully negating the quantifiers and using Lemma A.9, we conclude that there exist standard integers  $a_{*,j}$  for  $1 \leq j \leq l_*$  and  $a'_{i,j}$  for  $i = 1, 2, 3$  and  $1 \leq j \leq l'$ , not all zero, such that one has the relation

$$a_{*,1}\xi_{*,1} + \dots + a_{*,l_*}\xi_{*,l_*} + \sum_{i=1}^3 \sum_{j=1}^{l'} a'_{i,j}\xi'_{h_i,j} = O(\varepsilon) \pmod{1}$$

for many triples  $(h_1, h_2, h_3) \in (H')^3$ .

Suppose first that all of the  $a'_{i,j}$  vanish, so that we have a linear relation

$$a_{*,1}\xi_{*,1} + \dots + a_{*,l_*}\xi_{*,l_*} = O(\varepsilon) \pmod{1}$$

that only involves core frequencies. Then the situation is basically the same as that of Lemma 10.3; without loss of generality we may take  $a_{*,1} \neq 0$ , and if we then choose  $\tilde{\xi}_{*,2}, \dots, \tilde{\xi}_{*,l_*}$  so that  $a_{*,1}\tilde{\xi}_{*,j} = \xi_{*,j}$ , then we can rewrite

$$\xi_{*,1} = - \sum_{j=2}^{l'} a_{*,j} \tilde{\xi}_{*,j} + q + s \pmod{1}$$

for some  $q \in \mathbb{Q}$  and  $s = O(\varepsilon)$ , and one can then replace the  $\xi_{*,1}, \dots, \xi_{*,l_*}$  with  $\tilde{\xi}_{*,2}, \dots, \tilde{\xi}_{*,l_*}$  (decrementing  $l_*$  by 1) and append  $q$  and  $s$  to each of the collections  $\xi''_{h,1}, \dots, \xi''_{h,l''}$  and  $\xi'''_{h,1}, \dots, \xi'''_{h,l'''}$  respectively for each  $h \in H$ , contradicting minimality.

Now suppose that not all of the  $a'_{i,j}$  vanish; without loss of generality we may assume that  $a'_{1,1}$  is non-zero. By the pigeonhole principle, we can find  $h_2, h_3 \in H'$  such that

$$a_{*,1}\xi_{*,1} + \dots + a_{*,l_*}\xi_{*,l_*} + \sum_{i=1}^3 \sum_{j=1}^{l'} a'_{i,j} \xi'_{h_i,j} = O(\varepsilon) \pmod{1}$$

for all  $h_1$  in a dense subset  $H''$  of  $H'$ . Now let  $\tilde{\xi}_{*,j} \in {}^*\mathbb{T}$  for  $1 \leq j \leq l_*$  and  $\tilde{\xi}'_{h,j} \in {}^*\mathbb{T}$  for  $h_1 \in H'$  and  $1 \leq j \leq l'$  be such that  $a'_{1,1}\tilde{\xi}_{*,j} = \xi_{*,j}$  and  $a'_{1,1}\tilde{\xi}'_{h,j} = \xi'_{h,j}$ , then we have

$$\xi'_{h_1,1} = - \sum_{j=2}^{l'} a'_{1,j} \tilde{\xi}'_{h_1,j} - \sum_{j=1}^{l_*} a_{*,j} \tilde{\xi}_{*,j} - \sum_{i=2}^3 \sum_{j=1}^{l'} a'_{i,j} \tilde{\xi}'_{h_i,j} + q_{h_1} + s_{h_1} \pmod{O(1)}$$

for some standard rational  $q_{h_1}$  and some  $s_{h_1} = O(\varepsilon)$ . Furthermore one can easily ensure that  $q_{h_1}, s_{h_1}$  depend in a limit fashion on  $h_1$ . By Lemma A.9 (and refining  $H'$ ) we may assume that  $q_{h_1} = q_*$  is independent of  $h_1$ . We may thus replace  $H'$  by  $H''$  and replace  $\xi'_{h,1}, \dots, \xi'_{h,l'}$  by  $\tilde{\xi}'_{h,2}, \dots, \tilde{\xi}'_{h,l'}$  (decrementing  $l'$  by 1), while appending  $q_*$  and  $s_h$  to  $\xi''_{*,1}, \dots, \xi''_{*,l''}$  and  $\xi'''_{h,1}, \dots, \xi'''_{h,l'''}$  respectively, and replacing  $\xi_{*,1}, \dots, \xi_{*,l_*}$  by  $\tilde{\xi}_{*,1}, \dots, \tilde{\xi}_{*,l_*}, \tilde{\xi}'_{h_2,1}, \dots, \tilde{\xi}'_{h_2,l'}, \tilde{\xi}'_{h_3,1}, \dots, \tilde{\xi}'_{h_3,l'}$  (incrementing  $l_*$  as necessary). This contradicts the minimality of the partial solution, and the claim follows.  $\square$

This is still too simplistic for our applications, as the independence hypothesis on triples  $(h_1, h_2, h_3)$  will not quite be strong enough to give everything we need. Ideally, (in view of Proposition 7.3) we would like to have independence of the  $\xi_{*,1}, \dots, \xi_{*,l_*}, \xi'_{h_1,1}, \dots, \xi'_{h_4,l'}$  for almost all additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H'$ . Unfortunately, this need not be the case; indeed, if the original  $\xi_{h,i}$  are linear in  $h$ , say  $\xi_{h,i} = \alpha_i h$  for some  $\alpha_i \in {}^*\mathbb{T}$  and all  $1 \leq i \leq l'$ , then we have  $\xi_{h_1,i} + \xi_{h_2,i} = \xi_{h_3,i} + \xi_{h_4,i}$  for all additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H'$  and all  $1 \leq i \leq l'$ , and as a consequence it is not possible to obtain a decomposition as in Lemma 10.4 with the stronger independence property mentioned above. A similar obstruction occurs if the  $\xi_{h,i}$  are *bracket-linear* in  $h$ , for instance if  $\xi_{h,i} = \{\alpha_i h\} \beta_i \pmod{1}$  for some  $\alpha_i \in {}^*\mathbb{T}$  and  $\beta_i \in {}^*\mathbb{R}$ .

By using tools from additive combinatorics, we can show that bracket-linear frequencies are the *only* obstructions to independence on additive quadruples. More precisely, we have

**Lemma 10.5.** *Let  $l \in \mathbb{N}$ , let  $\varepsilon > 0$  be a limit real, let  $H$  be a dense limit subset of  $[[N]]$ , and for each  $h \in H$ , let  $\xi_{h,1}, \dots, \xi_{h,l}$  be frequencies in  ${}^*\mathbb{T}$  that depend in a limit fashion on  $h$ . Then there exists a dense subset  $H'$  of  $H$ , standard natural numbers,  $l_*, l', l'', l''', l'''' \in \mathbb{N}$ , “core” frequencies  $\xi_{*,1}, \dots, \xi_{*,l_*}, \xi''_{*,1}, \dots, \xi''_{*,l''} \in {}^*\mathbb{T}$ , and “petal” frequencies  $\xi'_{h,1}, \dots, \xi'_{h,l'}, \xi'''_{h,1}, \dots, \xi'''_{h,l'''}, \xi''''_{h,1}, \dots, \xi''''_{h,l''''} \in {}^*\mathbb{T}$  for each  $h \in H'$  depending in a limit fashion on  $h$ , obeying the following properties:*

- (i) (Independence) *For almost all additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H'$  (i.e. for all but  $o(|H'|^3)$  such quadruples), the frequencies  $\xi_{*,j}$  for  $1 \leq j \leq l_*$ ,  $\xi'_{h_i,j}$  for  $i = 1, 2, 3, 4$  and  $1 \leq j \leq l'$ , and  $\xi'''_{h_i,j}$  for  $i = 1, 2, 3$  and  $1 \leq j \leq l'''$  are jointly linearly independent modulo  $O(\varepsilon)$ .*
- (ii) (Rationality) *For each  $1 \leq j \leq l''$ ,  $\xi''_{*,j}$  is a standard rational.*
- (iii) (Smallness) *For each  $h \in H'$  and  $1 \leq j \leq l'''$ ,  $\xi'''_{h,j} = O(\varepsilon)$ .*
- (iv) (Bracket-linearity) *For each  $1 \leq j \leq l''''$ , there exist  $\alpha_j \in {}^*\mathbb{T}$  and  $\beta_j \in {}^*\mathbb{R}$  such that  $\xi''''_{h,j} = \{\alpha_j h\} \beta_j \bmod 1$  for all  $h \in H'$ . Furthermore, the map  $h \mapsto \xi''''_{h,j}$  is a Freiman homomorphism on  $H'$  (see §3 for the definition of a Freiman homomorphism).*
- (v) (Representation) *For each  $h \in H'$ , the  $\xi_{h,1}, \dots, \xi_{h,l}$  are linear combinations over  $\mathbb{Z}$  of the frequencies*

$$\xi_{*,1}, \dots, \xi_{*,l_*}, \xi'_{h,1}, \dots, \xi'_{h,l'}, \xi''_{*,1}, \dots, \xi''_{*,l''}, \xi'''_{h,1}, \dots, \xi'''_{h,l'''}, \xi''''_{h,1}, \dots, \xi''''_{h,l''''}.$$

*Proof.* As usual, we define a *partial solution* to be a collection of objects  $H'$ ,  $l_*, l', l'', l''', l''''$ ,  $\xi_{*,1}, \dots, \xi''''_{h,l''''}$ , obeying all of the required properties except possibly for the independence property. Again, there is clearly at least one partial solution, so we select a partial solution with a minimal value of  $l'$ , and then (for fixed  $l'$ ) a minimal value of  $l'''$ , and then (for fixed  $l', l'''$ ) a minimal value of  $l_*$ . We claim that this partial solution obeys the independence property, which will give the lemma.

Suppose for contradiction that this were not the case; then by Lemma A.9, there exist standard integers  $a_{*,j}$  for  $1 \leq j \leq l_*$ ,  $a'_{i,j}$  for  $1 \leq i \leq 4$  and  $1 \leq j \leq l'$ , and  $a''_{i,j}$  for  $1 \leq i \leq 3$  and  $1 \leq j \leq l'''$ , not all zero, such that

$$\sum_{j=1}^{l_*} a_{*,j} \xi_{*,j} + \sum_{i=1}^4 \sum_{j=1}^{l'} a'_{i,j} \xi'_{h_i,j} + \sum_{i=1}^3 \sum_{j=1}^{l'''} a''_{i,j} \xi'''_{h_i,j} = O(\varepsilon) \bmod 1$$

for many additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H'$ .

Suppose first that all the  $a'_{i,j}$  and  $a''_{i,j}$  vanished. Then we have a relation

$$\sum_{j=1}^{l_*} a_{*,j} \xi_{*,j} = O(\varepsilon) \bmod 1$$

that only involves core frequencies; arguing as in Lemma 10.4 we can thus find another partial solution with a smaller value of  $l_*$  (and the same value of  $l', l'''$ ), contradicting minimality.

Next, suppose that the  $a'_{i,j}$  all vanished, but the  $a''_{i,j}$  did not all vanish. Then we have a relation

$$\sum_{j=1}^{l_*} a_{*,j} \xi_{*,j} + \sum_{i=1}^3 \sum_{j=1}^{l'''} a''_{i,j} \xi'''_{h_i,j} = O(\varepsilon) \bmod 1 \quad (10.1)$$

for many triples  $h_1, h_2, h_3$  in  $H'$ .

Without loss of generality let us suppose that  $a_{1,1}''''$  is non-zero. By the pigeonhole principle, we may find  $h_2, h_3 \in H'$  such that (10.1) holds for all  $h_1$  in a dense subset  $H''$  of  $H'$ . As in previous arguments, we then find  $\tilde{\xi}_{*,j} \in {}^*\mathbb{T}$  such that  $a_{1,1}''''\tilde{\xi}_{*,j} = \xi_{*,j}$  for each  $1 \leq j \leq l_*$ , and also find  $\tilde{\beta}_j \in {}^*\mathbb{R}$  such that  $a_{1,1}''''\tilde{\beta}_j = \beta_j$  for all  $1 \leq j \leq l''''$ . If we then set  $\tilde{\xi}_{h,j}'''' := \{\alpha_j h\}\tilde{\beta}_j$  for each  $h \in H'$  and  $1 \leq j \leq l''''$ , then  $a_{1,1}''''\tilde{\xi}_{h,j}'''' = \xi_{h,j}''''$ , and so for any  $h_1 \in H'$  we have

$$\xi_{h_1,1}'''' = - \sum_{j=1}^{l_*} a_{*,j} \tilde{\xi}_{*,j} - \sum_{j=2}^{l''''} a_{1,j}'''' \tilde{\xi}_{h_1,j}'''' - \sum_{i=2}^3 \sum_{j=1}^{l''''} a_{i,j}'''' \tilde{\xi}_{h_i,j}'''' + q_{h_1} + s_{h_1} \pmod{1}$$

for some standard rational  $q_{h_1}$  and some  $s_{h_1} = O(\varepsilon)$ , both depending on a limit fashion on  $h_1$ . By refining  $H'$  if necessary (and using the bracket-linear nature of the  $\tilde{\xi}_{h,j}''''$ ) we may assume that the map  $h \mapsto \tilde{\xi}_{h,j}''''$  is a Freiman homomorphism on  $H'$ , and by Lemma A.9 we may make  $q_{h_1} = q_*$  independent of  $h_1$ . If we then argue as in the proof of Lemma 10.4, we may find a new partial solution with a smaller value of  $l''''$  and the same value of  $l'$ , contradicting minimality.

Finally, suppose that the  $a'_{i,j}$  did not all vanish. Using the Freiman homomorphism property to permute the  $i$  indices if necessary, we may assume that  $a'_{4,1}$  does not vanish. We then have

$$\Xi_1(h_1) + \Xi_2(h_2) + \Xi_3(h_3) + \Xi_4(h_4) = O(\varepsilon)$$

for many additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H'$ , where the limit functions  $\Xi_i : H \rightarrow {}^*\mathbb{T}$  are defined by

$$\Xi_i(h) := \sum_{j=1}^{l'} a'_{i,j} \xi'_{h,j} + \sum_{j=1}^{l''''} a_{i,j}'''' \xi_{h,j}'''' \pmod{1}$$

for  $i = 1, 2, 3$  and  $h \in H$ , and

$$\Xi_4(h) := \sum_{j=1}^{l_*} a_{*,j} \xi_{*,j} + \sum_{j=1}^{l'} a'_{4,j} \xi'_{h,j} \pmod{1}.$$

We can use this additive structure to “solve” for  $\Xi_4$ , using a result from additive combinatorics which we present here as Lemma F.1. Applying this lemma, we can then find a dense limit subset  $H'$  of  $H$ , a standard integer  $K$ , and frequencies  $\alpha'_1, \dots, \alpha'_K, \delta \in {}^*\mathbb{T}$  and  $\beta'_1, \dots, \beta'_K \in {}^*\mathbb{R}$  such that

$$\Xi_4(h) = \sum_{k=1}^K \{\alpha'_k h\} \beta'_k + \delta + O(\varepsilon) \pmod{1}$$

and thus

$$a'_{4,1} \xi'_{h,1} = \sum_{k=1}^K \{\alpha'_k h\} \beta'_k + \delta - \sum_{j=1}^{l_*} a_{*,j} \xi_{*,j} + \sum_{j=2}^{l'} a'_{4,j} \xi'_{h,j} + O(\varepsilon) \pmod{1}$$

for all  $h \in H'$ .

As usual, we now find  $\tilde{\beta}_k \in {}^*\mathbb{R}$  for  $1 \leq k \leq K$ ,  $\tilde{\beta}_j \in {}^*\mathbb{R}$  for  $1 \leq j \leq l''''$ ,  $\tilde{\delta} \in \mathbb{T}$  and  $\tilde{\xi}_{*,j}$  for  $1 \leq j \leq l_*$  such that  $a'_{4,1} \tilde{\beta}_k = \beta_k$ ,  $a'_{4,1} \tilde{\beta}_j = \beta_j$ ,  $a'_{4,1} \tilde{\delta} = \delta$ , and  $a'_{4,1} \tilde{\xi}_{*,j} = \xi_{*,j}$ .

We then set  $\tilde{\xi}'_{h,j} := \{\alpha_j h\} \tilde{\beta}_j \bmod 1$ , and we conclude that

$$\xi'_{h,1} = \sum_{k=1}^K \{\alpha'_k h\} \tilde{\beta}'_k + \tilde{\delta} - \sum_{j=1}^{l_*} a_{*,j} \tilde{\xi}_{*,j} + \sum_{j=2}^{l'} a'_{4,j} \tilde{\xi}'_{h,j} + q_h + s_h \bmod 1$$

for all  $h \in H'$ , where  $q_h \in \mathbb{Q}$  and  $s_h = O(\varepsilon)$  depend in a limit fashion on  $h$ . By refining  $H'$  we may take  $q_h = q_*$  independent of  $h$ .

We can then use relation to build a new partial solution that decreases  $l'$  by 1, at the expense of enlarging the other dimensions  $l_*, l'', l''', l''''$  (and also refining  $H$  to  $H'$ ), again contradicting minimality, and the claim follows.  $\square$

We now apply the above lemma to the language of horizontal frequency vectors introduced in the previous section. We need some definitions:

**Definition 10.6** (Properties of horizontal frequency vectors). Let

$$\mathcal{F} = (\xi_{i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_i} \quad \text{and} \quad \mathcal{F}' = (\xi'_{i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D'_i}$$

be horizontal frequency vectors.

- We say that  $\mathcal{F}$  is *independent* if, for each  $1 \leq i \leq d$ , the tuple  $(\xi_{i,j})_{1 \leq j \leq D_i}$  is independent modulo  $O(N^{-i})$ .
- We say that  $\mathcal{F}$  is *rational* if all the  $\xi_{i,j}$  are standard rationals.
- We say that  $\mathcal{F}$  is *small* if one has  $\xi_{i,j} = O(N^{-i})$  for all  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_i$ .
- We define the *disjoint union*  $\mathcal{F} \uplus \mathcal{F}' = (\xi''_{i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_i + D'_i}$  by declaring  $\xi''_{i,j}$  to equal  $\xi_{i,j}$  if  $j \leq D_i$  and  $\xi'_{i,j-D_i}$  if  $D_i < j \leq D_i + D'_i$ . This is clearly a horizontal frequency vector with dimensions  $(D_1 + D'_1, \dots, D_{s-1} + D'_{s-1})$ .
- We say that  $\mathcal{F}$  is *represented* by  $\mathcal{F}'$  if for every  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_i$ ,  $\xi_{i,j}$  is a standard integer linear combination of the  $\xi'_{i,j'}$  for  $1 \leq j' \leq D'_i$ .

**Lemma 10.7** (Sunflower lemma). *Let  $H$  be a dense subset of  $[[N]]$ , and let  $(\mathcal{F}_h)_{h \in H}$  be a family of horizontal frequency vectors depending in a limit fashion on  $h$ , whose dimension vector  $\vec{D} = \vec{D}_h$  is independent of  $h$ . Then we can find the following objects:*

- A dense subset  $H'$  of  $H$ ;
- Dimension vectors  $\vec{D}_* = \vec{D}_{*,\text{ind}} + \vec{D}_{*,\text{rat}}$  and  $\vec{D}' = \vec{D}'_{\text{lin}} + \vec{D}'_{\text{ind}} + \vec{D}'_{\text{sml}}$ , which we write as  $\vec{D}_* = (D_{*,i})_{i=1}^{s-1}$ ,  $\vec{D}_{*,\text{ind}} = (D_{*,\text{ind},i})_{i=1}^{s-1}$ , etc.;
- A core horizontal frequency vector  $\mathcal{F}_* = (\xi_{*,i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_{*,i}}$ , which is partitioned as  $\mathcal{F}_* = \mathcal{F}_{*,\text{ind}} \uplus \mathcal{F}_{*,\text{rat}}$ , with the indicated dimension vectors  $\vec{D}'_{\text{ind}}, \vec{D}'_{\text{rat}}$ ;
- A petal horizontal frequency vector  $\mathcal{F}'_h = (\xi'_{h,i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D'_i}$ , which is partitioned as  $\mathcal{F}'_h = \mathcal{F}'_{h,\text{lin}} \uplus \mathcal{F}'_{h,\text{ind}} \uplus \mathcal{F}'_{h,\text{sml}}$ , which is a limit function of  $h$  and with the indicated dimension vectors  $\vec{D}'_{\text{lin}}, \vec{D}'_{\text{ind}}, \vec{D}'_{\text{sml}}$

which obey the following properties:

- For all  $h \in H'$ ,  $\mathcal{F}'_{h,\text{sml}}$  are small.
- $\mathcal{F}_{*,\text{rat}}$  is rational.
- For every  $1 \leq i \leq d$  and  $1 \leq j \leq D'_{i,\text{lin}}$ , there exists  $\alpha_{i,j} \in {}^*\mathbb{T}$  and  $\beta_{i,j} \in {}^*\mathbb{R}$  such that (10.2) holds for all  $h \in H'$ , and furthermore that the map  $h \mapsto \xi'_{h,i,j}$  is a Freiman homomorphism on  $H'$ .
- For all  $h \in H$ ,  $\mathcal{F}_h$  is represented by  $\mathcal{F}_* \cup \mathcal{F}'_h$

- (Independence property) *For almost all additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ ,*

$$\mathcal{F}_{*,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}'_{h_i,\text{ind}} \uplus \biguplus_{i=1}^3 \mathcal{F}'_{h_i,\text{lin}}$$

*is independent.*

*Proof.* Write  $\mathcal{F}_h = (\xi_{h,i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_i}$ . For each  $1 \leq i \leq s-1$  in turn, apply Lemma 10.5 to the collections  $(\xi_{h,i,1}, \dots, \xi_{h,i,D_i})_{h \in H}$  and  $\varepsilon = O(N^{-i})$ , refining  $H$  once for each  $i$ . The claim then follows by relabeling.  $\square$

To apply this lemma to families of nilcharacters, we will need two additional lemmas.

**Lemma 10.8** (Change of basis). *Suppose that  $\chi \in \Xi_{\text{DR}}^{(s-1,r_*)}([N])$  is a degree-rank  $(s-1, r_*)$  nilcharacter with a total frequency representation  $(\vec{D}, \mathcal{F}, \eta)$ , and suppose that  $\mathcal{F}$  is represented by another horizontal frequency vector  $\mathcal{F}'$  with a dimension vector  $\vec{D}'$ . Then there exists a vertical frequency  $\eta' : G_{s-1}^{\vec{D}'} \rightarrow \mathbb{R}$  such that  $\chi$  has a total frequency representation  $(\vec{D}', \mathcal{F}', \eta')$ .*

*Proof.* By hypothesis, each element  $\xi_{i,j}$  of  $\mathcal{F}$  can be expressed as a standard linear combination  $\xi_{i,j} = \sum_{j'=1}^{D'_i} c_{i,j,j'} \xi'_{i,j'}$  of elements  $\xi'_{i,j'}$  of  $\mathcal{F}'$  of the same degree, where  $c_{i,j,j'} \in \mathbb{Z}$ .

Now let  $\psi : G^{\vec{D}'} \rightarrow G^{\vec{D}}$  be the unique filtered homomorphism that maps  $e'_{i,j'}$  to  $\prod_{j=1}^{D_i} e_{i,j}^{c_{i,j,j'}}$  (this can be viewed as an “adjoint” of the representation of  $\mathcal{F}$  by  $\mathcal{F}'$ ). By hypothesis,  $\chi$  has a representation  $\chi(n) = F(\mathcal{O}(n), \mathcal{O}_0(n))$  of  $\chi$  with

$$\text{Taylor}_i(\mathcal{O}) = \pi_{\text{Horiz}_i(G/\Gamma)} \left( \phi \left( \prod_{j=1}^{D_i} e_{i,j}^{\xi_{i,j}} \right) \right)$$

for some filtered homomorphism  $\phi : G^{\vec{D}} \rightarrow G$ . A brief calculation shows that the right-hand side can also be expressed as

$$\pi_{\text{Horiz}_i(G/\Gamma)} \left( \phi \circ \psi \left( \prod_{j=1}^{D'_i} (e'_{i,j})^{\xi'_{i,j}} \right) \right).$$

As  $\phi \circ \psi : G^{\vec{D}'} \rightarrow G$  is a filtered homomorphism, and  $\eta \circ \psi : G_{(s-1,r_*)}^{\vec{D}'} \rightarrow \mathbb{R}$  is a vertical frequency, we obtain the claim.  $\square$

**Lemma 10.9.** *Let  $\mathcal{F}$  be a horizontal frequency vector of dimension  $\vec{D}$  of the form*

$$\mathcal{F} = \mathcal{F}_{\text{rat}} \uplus \mathcal{F}_{\text{sml}} \uplus \mathcal{F}'$$

*where  $\mathcal{F}_{\text{rat}}$  is rational and  $\mathcal{F}_{\text{sml}}$  is small, and  $\mathcal{F}'$  has dimension  $\vec{D}'$ . Suppose that  $\chi \in \Xi_{\text{DR}}^{(s-1,r_*)}([N])$  is a nilcharacter with a total frequency representation  $(\vec{D}, \mathcal{F}, \eta)$ . Then there exists a vertical frequency  $\eta' : G_{s-1}^{\vec{D}'} \rightarrow \mathbb{R}$  such that  $\chi$  has total frequency  $(\vec{D}', \mathcal{F}'/M, \eta')$  for some standard integer  $M \geq 1$ .*

*Remark.* This lemma crucially relies on the hypothesis  $s \geq 3$ , as it makes the (degree 1) contributions of rational and small frequencies to be of lower order. Because of this, the inverse conjecture for  $s > 2$  is in a very slight way a little bit

simpler than the  $s \leq 2$  theory, though it is of course more complicated in many other ways.

*Proof.* By induction we may assume that  $\mathcal{F}$  is formed from  $\mathcal{F}'$  by adding a single frequency  $\xi_{i_0, D_{i_0}}$ , which is either rational or small.

Let us first suppose that we are adding a single frequency which is not just rational, but is in fact an integer. Then if  $\chi(n) = F(g(n)^*\Gamma, g_0(n)^*\Gamma_0)$  is a nilcharacter with a total frequency representation  $(\vec{D}, \mathcal{F}, \eta)$ , then we have a filtered homomorphism  $\phi : G^{\vec{D}}/\Gamma^{\vec{D}} \rightarrow G/\Gamma$  such that

$$g_i = \prod_{j=1}^{D_i} \phi(e_{i,j})^{\xi_{i,j}} \bmod G_{(i,1)}$$

for all  $1 \leq i \leq s_* - 1$ , where  $g_i$  are the Taylor coefficients of  $g$ . Specialising to the degree  $i_0$  and using the integer nature of  $\xi_{i_0, D_{i_0}}$ , we have

$$g_{i_0} = g'_{i_0} \gamma_{i_0}$$

where  $\gamma_{i_0}$  is an element of  $\Gamma_{i_0}$ , and

$$g'_i = \prod_{j=1}^{D_i-1} \phi(e_{i,j})^{\xi_{i,j}} \bmod G_{(i,1)}.$$

From this and the Baker-Campbell-Hausdorff formula (3.2), we can write  $g(n) = g'(n)\gamma_{i_0}^{\binom{n}{i_0}}$ , where  $g'$  is a polynomial sequence with a horizontal frequency representation  $(\vec{D}', \phi', \mathcal{F}')$ , where  $\vec{D}'$  is  $\vec{D}$  with  $D_{i_0}$  decremented by one, and  $\phi'$  is the restriction of  $\phi$  to the subnilmanifold  $G^{\vec{D}'}/\Gamma^{\vec{D}'}$ . Since  $g(n)^*\Gamma = g'(n)^*\Gamma$ , we see that  $\chi$  has a total frequency representation  $(\vec{D}', \mathcal{F}', \eta')$ , where  $\eta'$  is the restriction of  $\eta : G^{\vec{D}}_{(s-1, r_*)} \rightarrow \mathbb{R}$  to  $G^{\vec{D}'}_{(s-1, r_*)}$ . This gives the claim in this case (with  $M = 1$ ).

Now suppose that  $\xi_{i_0, D_{i_0}}$  is merely rational rather than integer. Then we can argue as before, except that now  $\gamma_{i_0}$  is a rational element of  $G_{i_0}$ , so that  $\gamma_{i_0}^m \in \Gamma_{i_0}$  for some standard positive integer  $m$ . As such, there exists a standard positive integer  $q$  such that  $\gamma_{i_0}^{\binom{n}{i_0}} \bmod * \Gamma$  is periodic with period  $q$ . As a consequence, there exists a bounded index subgroup  $\Gamma'$  of  $\Gamma$  such that the point

$$g'(n)\gamma_{i_0}^{\binom{n}{i_0}} \bmod * \Gamma$$

in  $G/\Gamma$  can be expressed as a Lipschitz function of

$$g'(n) \bmod * \Gamma'$$

and of the quantity  $n/q \bmod 1$ . Repeating the previous arguments, we thus obtain a total frequency representation  $(\vec{D}', \tilde{\mathcal{F}}', \eta')$  for some  $\eta'$ , and some  $\tilde{\mathcal{F}}'$  whose coefficients are rational combinations of those of  $\mathcal{F}'$ ; note that the  $n/q$  dependence can be easily absorbed into the lower order term  $G_0/\Gamma_0$  since  $s \geq 3$ . The claim then follows from Lemma 10.8.

Finally, suppose that  $\xi_{i_0, D_{i_0}}$  is small rather than rational. Then we can write

$$g_{i_0} = c_{i_0} g'_{i_0}$$



where  $g'_{i_0}$  is as before, and  $c_{i_0} \in G_{i_0}$  is at a distance  $O(N^{-i_0})$  from the origin. We can thus write

$$g(n) = c_{i_0}^{(n)} g'(n)$$

where  $g'$  is a polynomial sequence with horizontal frequency representation

$$(\vec{D}', \phi', \mathcal{F}').$$

On  $[N]$ , the sequence  $c_{i_0}^{(n)}$  can be expressed as a bounded Lipschitz function of  $n/2N \bmod 1$ . As a consequence, we can thus write  $\chi$  in the form

$$\chi(n) = F'(g'(n) * \Gamma, g_0(n) * \Gamma_0, n/2N \bmod 1)$$

for some  $F' \in \text{Lip}^*(G/\Gamma \times G_0/\Gamma_0 \times \mathbb{T})$ . As  $s \geq 3$ , the final term  $\mathbb{T}$  can be absorbed into the degree-rank  $\leq (s-1, r_*-1)$  nilmanifold  $G_0/\Gamma_0$ , and the claim follows (with  $M=1$ ).  $\square$

Finally, we can state the main result of this section.

**Lemma 10.10** (Sunflower lemma). *Let  $H$  be a dense subset of  $[N]$ , and let  $(\chi_h)_{h \in H}$  be a family of nilcharacters  $\chi_h \in \Xi_{\text{DR}}^{(s-1, r_*)}([N])$  depending in a limit fashion on  $H$ . Then we can find*

- (i) A dense subset  $H'$  of  $H$ ;
- (ii) Dimension vectors  $\vec{D}_*$  and  $\vec{D}' = \vec{D}'_{\text{lin}} + \vec{D}'_{\text{ind}}$ , which we write as  $\vec{D}_* = (D_{*,i})_{i=1}^{s-1}$ ,  $\vec{D}' = (D'_i)_{i=1}^{s-1}$ ,  $\vec{D}'_{\text{lin}} = (D'_{\text{lin},i})_{i=1}^{s-1}$ ,  $\vec{D}'_{\text{ind}} = (D'_{\text{ind},i})_{i=1}^{s-1}$ ;
- (iii) A core horizontal frequency vector  $\mathcal{F}_* = (\xi_{*,i,j})_{1 \leq i \leq d; 1 \leq j \leq D_{*,i}}$ ;
- (iv) A petal horizontal frequency vector  $\mathcal{F}'_h = (\xi'_{h,i,j})_{1 \leq i \leq d; 1 \leq j \leq D'_{i,\text{lin}}}$ , which is partitioned as  $\mathcal{F}'_h = \mathcal{F}'_{h,\text{lin}} \uplus \mathcal{F}'_{h,\text{ind}}$ , which is a limit function of  $h$ , where  $\mathcal{F}'_{h,\text{lin}}, \mathcal{F}'_{h,\text{ind}}$  have dimensions  $\vec{D}'_{\text{lin}}, \vec{D}'_{\text{ind}}$  respectively;
- (v) A vertical frequency  $\eta : G_{(s-1, r_*)}^{\vec{D}_* + \vec{D}'} \rightarrow \mathbb{R}$  with dimension vector  $\vec{D}_* + \vec{D}'$

which obey the following properties:

- (i) ( $\mathcal{F}'_{h,\text{lin}}$  is bracket-linear) For every  $1 \leq i \leq d$  and  $1 \leq j \leq D'_{i,\text{lin}}$ , there exists  $\alpha_{i,j} \in {}^*\mathbb{T}$  and  $\beta_{i,j} \in {}^*\mathbb{R}$  such that

$$\xi'_{h,i,j} = \{\alpha_{i,j} h\} \beta_{i,j} \bmod 1 \quad (10.2)$$

for all  $h \in H'$ , and furthermore that the map  $h \mapsto \xi'_{h,i,j}$  is a Freiman homomorphism on  $H'$ .

- (ii) (Independence) For almost all additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ ,

$$\mathcal{F}_{*,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}'_{h_i,\text{ind}} \uplus \biguplus_{i=1}^3 \mathcal{F}'_{h_i,\text{lin}}$$

is independent.

- (iii) (Representation) For all  $h \in H'$ ,  $\chi_h$  has a total frequency representation  $(\vec{D}_* + \vec{D}', \mathcal{F}_* \cup \mathcal{F}'_h, \eta)$ .

*Proof.* Each  $\chi_h$  thus has a total frequency representation  $(\vec{D}_h, \mathcal{F}_h, \eta_h)$ . The space of representations is a  $\sigma$ -limit set, so by Lemma A.11 we may assume that  $(\vec{D}_h, \mathcal{F}_h, \eta_h)$  depends in a limit fashion on  $h$ .

The number of possible dimension vectors is countable. Applying Lemma A.9, and passing from  $H$  to a dense subset, we may assume that  $\vec{D} = \vec{D}_h$  is independent of  $h$ .

We then apply Lemma 10.7 to the  $(\mathcal{F}_h)_{h \in H}$ , obtaining a dense subset  $H'$  of  $H$ , dimension vectors  $\vec{D}_* = \vec{D}_{*,\text{ind}} + \vec{D}_{*,\text{rat}}$  and  $\vec{D}' = \vec{D}'_{\text{lin}} + \vec{D}'_{\text{ind}} + \vec{D}'_{\text{sml}}$ , a core horizontal frequency vector  $\mathcal{F}_* = \mathcal{F}_{*,\text{ind}} \uplus \mathcal{F}_{*,\text{rat}}$ , and petal horizontal frequency vectors  $\mathcal{F}'_h = \mathcal{F}'_{h,\text{lin}} \uplus \mathcal{F}'_{h,\text{ind}} \uplus \mathcal{F}'_{h,\text{sml}}$  for each  $h \in H'$  with the stated properties.

Applying Lemma 10.8, we see that for each  $h \in H'$ ,  $\chi_h$  has a total frequency representation

$$(\vec{D}_* + \vec{D}', \mathcal{F}_* \uplus \mathcal{F}'_h, \eta'_h)$$

for some vertical frequency  $\eta'_h$ . Applying Lemma 10.9, we conclude that  $\chi_h$  has a total frequency representation

$$(\vec{D}_{*,\text{ind}} + \vec{D}'_{\text{lin}} + \vec{D}'_{\text{ind}}, \mathcal{F}_{*,\text{ind}} \uplus \mathcal{F}'_{h,\text{lin}} \uplus \mathcal{F}'_{h,\text{ind}}, \eta''_h)$$

for some vertical frequency  $\eta''_h$ . The number of vertical frequencies  $\eta''_h$  is countable, so by Lemma A.9 we may assume that  $\eta = \eta''_h$  is also independent of  $h$ . The claim then follows.  $\square$

## 11. OBTAINING BRACKET-LINEAR BEHAVIOUR

We return now to the task of proving Theorem 7.2. To recall the situation thus far, we are given a two-dimensional nilcharacter  $\chi \in \Xi_{\text{Multi}}^{(1,s-1)}(*\mathbb{Z}^2)$  and a family of degree-rank  $(s-1, r_*)$  nilcharacters  $(\chi_h)_{h \in H}$  depending in a limit fashion on a parameter  $h$  in a dense subset  $H$  of  $[[N]]$ , with the property that there is a function  $f \in L^\infty[N]$  such that  $\chi(h, \cdot) \otimes \chi_h$   $(s-2)$ -correlates with  $f$  for all  $h \in H$ . Using Proposition 7.3 to eliminate  $f$  and  $\chi$ , and refining  $H$  to a dense subset if necessary, we conclude that the nilcharacter (7.2) is  $(s-2)$ -biased for many additive quadruples  $h_1 + h_2 = h_3 + h_4$  in  $H$ . We make the simple but important remark that this conclusion is “hereditary” in the sense that it continues to hold if we replace  $H$  with an arbitrary dense subset  $H'$  of  $H$ , since the hypothesis of Proposition 7.3 clearly restricts from  $H$  to  $H'$  in this fashion.

Next, we apply Lemma 10.10 to obtain a dense refinement  $H'$  on  $H$  for which the  $\chi_h$  have a frequency representation involving various types of frequencies: a core set of frequencies  $\mathcal{F}_*$ , a bracket-linear family  $(\mathcal{F}'_{h,\text{lin}})_{h \in H'}$  of petal frequencies and an independent family  $(\mathcal{F}'_{h,\text{ind}})_{h \in H'}$  of petal frequencies.

iiiiiii .mine The main result of this section uses the bias of (7.2), combined with the quantitative equidistribution theory on nilmanifolds (as reviewed in Appendix D) to obtain an important milestone towards establishing Theorem 7.2, namely that the independent petal frequencies  $\mathcal{F}'_{h,\text{ind}}$  do not actually have any influence on the top-order behaviour of the nilcharacters  $\chi_h$ , and that the bracket-linear frequencies only influence this top-order behaviour in a linear fashion. For this, we use an argument of Furstenberg and Weiss [14], also used in the predecessor [28] to this paper. See also [29] for another exposition of this argument. ===== The main result of this section uses the bias of (7.2), combined with the quantitative equidistribution theory on nilmanifolds (as reviewed in Appendix D) to obtain an important milestone towards establishing Theorem 7.2, namely that the independent petal frequencies  $\mathcal{F}'_{h,\text{ind}}$  do not actually have any influence on the top-order behaviour of the nilcharacters  $\chi_h$ , and that the bracket-linear frequencies only influence this top-order behaviour in a linear fashion. For this, we use an argument of Furstenberg and Weiss [14] that was also used in the predecessor [28] to this paper. See [29] for another, somewhat simplified, exposition of this argument. llllllll .r207

We begin by formally stating the result we will prove in this section.

**Theorem 11.1** (No petal-petal or regular terms). *Let  $f, H, \chi, (\chi_h)_{h \in H}$  be as in Theorem 7.2 and let  $H', \vec{D}_*, \vec{D}', \vec{D}'_{\text{lin}}, \vec{D}'_{\text{ind}}, \mathcal{F}_*, \mathcal{F}'_h, \mathcal{F}'_{h,\text{lin}}, \mathcal{F}'_{h,\text{ind}}, \eta$  be as in Lemma 10.10. Let  $w \in G^{\vec{D}_* + \vec{D}'}$  be an  $r_* - 1$ -fold commutator of  $e_{i_1, j_1}, \dots, e_{i_{r_*}, j_{r_*}}$ , where  $1 \leq i_1, \dots, i_{r_*} \leq s-1$ ,  $i_1 + \dots + i_{r_*} = s-1$ , and  $1 \leq j_l \leq D_{*, i_l} + D'_{i_l}$  for all  $l$  with  $1 \leq l \leq r_*$ .*

- (i) (No petal-petal terms) *If  $j_l > D_{*, i_l}$  for at least two values of  $l$ , then  $\eta(w) = 0$ .*
- (ii) (No regular terms) *If  $j_l > D_{*, i_l} + D'_{\text{lin}, i_l}$  for at least one value of  $l$ , then  $\eta(w) = 0$ .*
- (iii) (No petal-petal terms) *If  $j_l > D_{*, i_l}$  for at least two values of  $l$  then  $\eta(w) = 0$ .*
- (iv) (No regular terms) *If  $j_l > D_{*, i_l} + D'_{\text{lin}, i_l}$  for at least one value of  $l$  then  $\eta(w) = 0$ .*

The remainder of this section is devoted to the proof of Theorem 11.1.

Let the notation and assumptions be as in the above theorem. From Proposition 7.3 we know that, for many additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H'$ , the sequence (7.2) is  $(s-2)$ -biased. Also, from Lemma 10.10, we see that for almost all of these quadruples, the horizontal frequency vectors

$$\mathcal{F}_{*, \text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}_{h_i, \text{ind}} \uplus \biguplus_{i=a, b, c} \mathcal{F}_{h_i, \text{lin}} \quad (11.1)$$

are independent for all distinct  $a, b, c \in \{1, 2, 3, 4\}$ . We may therefore find an additive quadruple  $(h_1, h_2, h_3, h_4)$  for which (7.2) is  $(s-2)$ -biased, and for which (11.1) is independent for all choices of distinct  $a, b, c \in \{1, 2, 3, 4\}$ .

Fix  $(h_1, h_2, h_3, h_4)$  with these properties. We convert the above information to a non-equidistribution result concerning a polynomial orbit.

For each  $i = 1, 2, 3, 4$ , we see from Lemma 10.10 that  $\chi_{h_i}$  has a total frequency representation

$$(\vec{D}_* + \vec{D}', \mathcal{F}_* \uplus \mathcal{F}'_{h_i}, \eta).$$

We write

$$\mathcal{F}_* \uplus \mathcal{F}'_{h_i} = (\xi_{h_i, j, k})_{1 \leq j \leq s-1; 1 \leq k \leq D_j},$$

where

$$D_j = D_{*, j} + D'_j;$$

thus the frequencies associated to  $\mathcal{F}_*, \mathcal{F}'_{h_i, \text{ind}}, \mathcal{F}'_{h_i, \text{lin}}$  correspond to the ranges  $1 \leq k \leq D_{*, j}$ ,  $D_{*, j} < k \leq D_{*, j} + D'_{\text{ind}, j}$ , and  $D_{*, j} + D'_{\text{ind}, j} < k \leq D_j$  respectively.

As (7.2) is  $(s-2)$ -biased, we conclude that

$$|\mathbb{E}_{n \in [N]} \chi_{h_1}(n) \otimes \chi_{h_2}(n + h_1 - h_4) \otimes \overline{\chi_{h_3}}(n) \otimes \overline{\chi_{h_4}}(n + h_1 - h_4) \psi_{h_1, h_2, h_3, h_4}(n)| \gg 1 \quad (11.2)$$

for some degree  $\leq (s-2)$  nilsequence  $\psi_{h_1, h_2, h_3, h_4}$ , where  $\chi_h$  is defined to be zero outside of  $[N]$ . As any cutoff to an interval can be approximated to arbitrary standard accuracy by a degree 1 nilsequence, and  $s \geq 3$ , we see that the same claim holds if  $\chi_h$  is instead extended to be a nilsequence on all of  $^*\mathbb{Z}$ .

From Definition 6.1 and the total frequency representation of the  $\chi_{h_i}$ , we can rewrite the sequence inside the expectation of (11.2) as a degree-rank  $\leq (s-1, r_*)$  nilsequence  $n \mapsto F(\mathcal{O}(n))$ . Here  $G/\Gamma$  is the product nilmanifold<sup>6</sup>

$$G/\Gamma := \left( \prod_{i=1}^4 G_{(i)}/\Gamma_{(i)} \right) \times G_{(0)}/\Gamma_{(0)}$$

for some filtered nilmanifold  $G_{(0)}/\Gamma_{(0)}$  of degree-rank  $< (s-1, r_*-1)$  and filtered nilmanifolds  $G_{(i)}/\Gamma_{(i)}$  of degree-rank  $\leq (s-1, r_*)$  for  $i = 1, 2, 3, 4$ . The orbit  $\mathcal{O}$  is defined by

$$\mathcal{O} = (\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \mathcal{O}_4, \mathcal{O}_0) \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$$

where, for each  $i, j$  with  $1 \leq i \leq 4$  and  $1 \leq j \leq s-1$  we have

$$\text{Taylor}_j(\mathcal{O}_{(i)}) = \pi_{\text{Horiz}_j(G_{(i)}/\Gamma_{(i)})} \left( \phi_{(i)} \left( \prod_{1 \leq k \leq D_j} e_{j,k}^{\xi_{h_i,j,k}} \right) \right) \quad (11.3)$$

where  $\vec{D} := (D_1, \dots, D_{s-1})$ ,  $\phi_{(i)} : G^{\vec{D}}/\Gamma^{\vec{D}} \rightarrow G_{(i)}/\Gamma_{(i)}$  is a filtered homomorphism and  $\pi_{\text{Horiz}_j(G_{(i)}/\Gamma_{(i)})} : (G_{(i)})_j \rightarrow \text{Horiz}_j(G_{(i)}/\Gamma_{(i)})$  is the projection to the  $j^{\text{th}}$  horizontal torus. Finally  $F \in \text{Lip}({}^*(G/\Gamma))$  is defined by

$$F(\phi_{(1)}(t_{(1)})x_{(1)}, \dots, \phi_{(4)}(t_{(4)})x_{(4)}, y) = e((\eta(t_{(1)}) + \eta(t_{(2)}) - \eta(t_{(3)}) - \eta(t_{(4)}))) F(x_{(1)}, \dots, x_{(4)}, y) \quad (11.4)$$

for all  $(x_{(1)}, \dots, x_{(4)}, y) \in G/\Gamma$  and  $t_{(1)}, \dots, t_{(4)} \in G_{(s-1, r_*)}^{\vec{D}}$ . (Note that the shifts by  $h_1 - h_4$  in (11.2) do not affect the Taylor coefficients of  $\mathcal{O}_{(i)}$ , thanks to the remarks following Definition 9.6.)

By hypothesis, we have

$$|\mathbb{E}_{n \in [N]} F(\mathcal{O}(n))| \gg 1.$$

Applying Theorem D.6, we conclude that

$$|\int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x)| \gg 1 \quad (11.5)$$

for some bounded  $\varepsilon \in G$  and some rational subgroup  $G_P$  of  $G$  with the property that

$$\pi_{\text{Horiz}_j(G)}(G_P \cap G_{(i)}) \geq \Xi_j^\perp \quad (11.6)$$

for all  $1 \leq j \leq s-1$ , where

$$\Xi_j^\perp := \{x \in \text{Horiz}_j(G) : \xi_j(x) = 0 \text{ for all } \xi_j \in \Xi_j\}$$

and  $\Xi_j \leq \widehat{\text{Horiz}_j(G/\Gamma)}$  is the group of all (standard) continuous homomorphisms  $\xi_j : \text{Horiz}_j(G/\Gamma) \rightarrow \mathbb{T}$  such that

$$\xi_j(\text{Taylor}_j(\mathcal{O})) = O(N^{-j}).$$

From (11.4) and (11.5) we conclude the following lemma.

---

<sup>6</sup>Unfortunately, there will be several types of subscripts on nilpotent Lie groups  $G$  in this argument. Firstly one has the factor groups  $G_{(i)}$ . Then one also has the degree filtration groups  $G_d$  and the degree-rank filtration groups  $G_{(d,r)}$  of  $G$  (and also the analogous subgroups  $(G_{(i)})_d$ ,  $(G_{(i)})_{(d,r)}$  of the factor groups  $G_{(i)}$ ), as well as the free nilpotent groups  $G^{\vec{D}} = G_{(s-1, r_*)}^{\vec{D}}$ . Finally, a Ratner subgroup  $G_P$  of  $G$  will also make an appearance later. We hope that these notations can be kept separate from each other.

**Lemma 11.2.** *The group  $G_P \cap ((G_{(1)})_{(s-1, r_*)} \times \{\text{id}\} \times \{\text{id}\} \times \{\text{id}\} \times \{\text{id}\})$  is annihilated by  $\eta$ .*

*Proof.* Let  $g = (g_{(1)}, \text{id}, \text{id}, \text{id}, \text{id})$  lie in the indicated group. Then  $g$  is central, and so from the invariance of Haar measure we have

$$\int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x) = \int_{G_P/\Gamma_P} F(g\varepsilon x) d\mu(x).$$

On the other hand, from (11.4) we have

$$\int_{G_P/\Gamma_P} F(g\varepsilon x) d\mu(x) = e(\eta(g)) \int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x).$$

Comparing these relationships with (11.5) we obtain the claim.  $\square$

We now analyse the group  $G_P$  further. For each  $1 \leq j \leq s-1$ , let  $V_{123,j}$  denote the subgroup of  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)})$  generated by the diagonal elements

$$(\phi_{(1)}(e_{j,k}), \phi_{(2)}(e_{j,k}), \phi_{(3)}(e_{j,k}))$$

for  $1 \leq k \leq D_{*,j}$ , and by the elements

$$(\phi_{(1)}(e_{j,k}), \text{id}, \text{id}), (\text{id}, \phi_{(2)}(e_{j,k}), \text{id}), (\text{id}, \text{id}, \phi_{(3)}(e_{j,k}))$$

for  $D_{*,j} < k \leq D_j$ . We define the subgroup  $V_{124,j}$  of  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(4)})$  similarly by replacing (3) with (4) throughout.

**Lemma 11.3** (Components of  $G_P$ ). *Let  $1 \leq j \leq s-1$ . Then the projection of  $G_P \cap G_j$  to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)})$  contains  $V_{123,j}$ . Similarly, the projection to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(4)})$  contains  $V_{124,j}$ .*

*Proof.* We shall just prove the first claim; the second claim is similar (but uses  $\{a, b, c\} = \{1, 2, 4\}$  instead of  $\{a, b, c\} = \{1, 2, 3\}$ ).

Suppose the claim failed for some  $j$ . Using (11.6) and duality, we conclude that there exists a  $\xi_j \in \Xi_j$  which annihilates the kernel of the projection to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)})$ , and which is non-trivial on  $V_{123,j}$ . As  $\xi_j$  annihilates the kernel of the projection to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)})$ , we have a decomposition of the form

$$\xi_j(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(0)}) = \xi_{(1),j}(x_{(1)}) + \xi_{(2),j}(x_{(2)}) + \xi_{(3),j}(x_{(3)})$$

for  $x_{(i)} \in \text{Horiz}_j(G_{(i)})$  for  $i = 1, 2, 3, 4, 0$ , where  $\xi_{(i),j} : \text{Horiz}_j(G_{(i)}) \rightarrow \mathbb{R}$  for  $i = 1, 2, 3$  are characters.

By definition of  $\Xi_j$ , we conclude that

$$\xi_{(1),j}(\text{Taylor}_j(\mathcal{O}_{(1)})) + \xi_{(2),j}(\text{Taylor}_j(\mathcal{O}_{(2)})) + \xi_{(3),j}(\text{Taylor}_j(\mathcal{O}_{(3)})) = O(N^{-j}).$$

However, from (11.3) we have

$$\xi_{(i),j}(\text{Taylor}_j(\mathcal{O}_{(i)})) = \sum_{k=1}^{D_j} c_{(i),j,k} \xi_{h_i,j,k} \quad (11.7)$$

where the  $c_{(i),j,k}$  are standard integers, defined by the formula

$$c_{(i),j,k} := \xi_{(i),j}(\phi_{(i)}(e_{j,k})). \quad (11.8)$$

From the independence of (11.1) with  $\{a, b, c\} = \{1, 2, 3\}$ , we conclude that the  $c_{(i),j,k}$  all vanish for  $i = 1, 2, 3$  and  $D_{*,j} < k \leq D_j$ , and that the sum  $c_{(1),j,k} +$

$c_{(2),j,k} + c_{(3),j,k}$  vanishes for  $1 \leq k \leq D_{*,j}$ . But this forces  $\xi_j$  to vanish on  $V_{123,j}$ , contradiction.  $\square$

We now take commutators in the spirit of an argument of Furstenberg and Weiss [14] (see also [39, 50] for similar arguments in completely different settings) to conclude the following result which roughly speaking asserts that all “petal-petal interactions” are trivial.

**Corollary 11.4** (Furstenberg-Weiss commutator argument). *Let  $w$  be an  $r_* - 1$ -fold iterated commutator of generators  $e_{j_1,k_1}, \dots, e_{j_{r_*},k_{r_*}}$  with  $1 \leq j_l \leq s-1$ ,  $1 \leq k_l \leq D_l$  for  $l = 1, \dots, r_*$  and  $j_1 + \dots + j_{r_*} = s-1$  (thus  $w$  has “degree-rank  $(s-1, r_*)$ ” in some sense). Suppose that at least two of the generators, say  $e_{j_1,k_1}, e_{j_2,k_2}$ , are “petal” generators in the sense that  $k_1 > D_{*,j_1}$  and  $k_2 > D_{*,j_2}$ . Then  $(\phi_{(1)}(w), \text{id}, \text{id}, \text{id}, \text{id}) \in G_P$ .*

*Proof.* For  $e_{j_1,k_1}$ , we may invoke Lemma 11.3 and find an element  $g_{j_1,k_1}$  of  $G_P \cap G_{j_1}$  for which the coordinates 1, 2, 3 are equal (modulo projection to

$$\text{Horiz}_{j_1}(G_{(1)}) \times \text{Horiz}_{j_1}(G_{(2)}) \times \text{Horiz}_{j_1}(G_{(3)}))$$

to  $(\phi_1(e_{j_1,k_1}), \text{id}, \text{id})$ . Similarly, we may find an element  $g'_{j_2,k_2}$  of  $G_P \cap G_{j_2}$  for which the coordinates 1, 2, 4 are equal (modulo projection to

$$\text{Horiz}_{j_2}(G_{(1)}) \times \text{Horiz}_{j_2}(G_{(2)}) \times \text{Horiz}_{j_2}(G_{(4)}))$$

to  $(\phi_1(e_{j_2,k_2}), \text{id}, \text{id})$ . Finally, for all of the other  $e_{j,k}$ , we can find elements  $g''_{j,k}$  of  $G_P \cap G_j$  for which the first coordinate is equal (modulo projection to  $\text{Horiz}_j(G_{(1)})$ ) to  $\phi_{(1)}(e_{j,k})$ . If one then takes iterated commutators of the  $g_{j_1,k_1}, g'_{j_2,k_2}, g''_{j,k}$  in the order indicated by  $w$ , we see (using the filtration property, the homomorphism property of  $\phi_{(1)}$ , and the fact that the  $G_i/\Gamma_i$  have degree  $\leq (s-1, r_*)$  for  $i = 1, 2, 3, 4$  and degree  $< (s-1, r_*-1)$  for  $i = 0$ ) that we obtain the element  $(\phi_{(1)}(w), \text{id}, \text{id}, \text{id}, \text{id})$ . Since the iterated commutator of elements in  $G_P$  stays in  $G_P$ , the claim follows.  $\square$

From Lemma 11.2 and Corollary 11.4 we immediately obtain the first part (i) of Theorem 11.1. We now turn to the second part of the theorem. For this, we need two further variants of Lemma 11.3. For any  $1 \leq j \leq s-1$ , let  $V_{\text{ind},j}$  be the subspace of  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)}) \times \text{Horiz}_j(G_{(4)})$  generated by the elements

$$(\phi_{(1)}(e_{j,k}), \phi_{(2)}(e_{j,k}), \phi_{(3)}(e_{j,k}), \phi_{(4)}(e_{j,k}))$$

for  $1 \leq k \leq D_{*,j}$  and the elements

$$(\phi_{(1)}(e_{j,k}), \text{id}, \text{id}, \text{id}), (\text{id}, \phi_{(2)}(e_{j,k}), \text{id}, \text{id}), (\text{id}, \text{id}, \phi_{(3)}(e_{j,k}), \text{id}), (\text{id}, \text{id}, \text{id}, \phi_{(4)}(e_{j,k}))$$

for  $D_{*,j} < k \leq D_{*,j} + D'_{\text{ind},j}$ .

**Lemma 11.5** (Components of  $G_P$ , II). *Let  $1 \leq j \leq s-1$ . Then the projection of  $G_P \cap G_j$  to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)}) \times \text{Horiz}_j(G_{(4)})$  contains  $V_{\text{ind},j}$ .*

*Proof.* Suppose the claim failed for some  $j$ . Using (11.6) and duality, we conclude that there exists a  $\xi_j \in \Xi_j$  which annihilates the kernel of the projection to  $\text{Horiz}_j(G_{(1)}) \times \text{Horiz}_j(G_{(2)}) \times \text{Horiz}_j(G_{(3)}) \times \text{Horiz}_j(G_{(4)})$ , and which is non-trivial on  $V_{\text{ind},j}$ . In particular, we have a decomposition of the form

$$\xi_j(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(0)}) = \sum_{i=1}^4 \xi_{(i),j}(x_{(i)}) \quad (11.9)$$

for  $x_{(i)} \in \text{Horiz}_j(G_{(i)})$  for  $i = 1, 2, 3, 4, 0$ , where  $\xi_{(i),j} : \text{Horiz}_j(G_{(i)}) \rightarrow \mathbb{R}$  for  $i = 1, 2, 3, 4$  are characters.

By definition of  $\Xi_j$ , we conclude that

$$\sum_{i=1}^4 \xi_{(i),j}(\text{Taylor}_j(\mathcal{O}_{(i)})) = O(N^{-j}).$$

Inserting (11.7), we conclude that

$$\sum_{k=1}^{D_j} \sum_{i=1}^4 c_{(i),j,k} \xi_{h_i,j,k} = O(N^{-j}). \quad (11.10)$$

The left-hand side is an integer linear combination of the degree  $j$  frequencies in

$$\mathcal{F}_{*,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}_{h_i,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}_{h_i,\text{lin}}.$$

Using the Freiman homomorphism property from Lemma 10.10 we can eliminate the role of  $\mathcal{F}_{h_4,\text{lin}}$ , leaving only

$$\mathcal{F}_{*,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}_{h_i,\text{ind}} \uplus \biguplus_{i=1}^3 \mathcal{F}_{h_i,\text{lin}}.$$

But this is just (11.1) for  $\{a, b, c\} = \{1, 2, 3\}$ . We conclude that the coefficients of the left-hand side of (11.10) in this basis vanish, which in terms of the original coefficients  $c_{(i),j,k}$  means that

$$\sum_{i=1}^4 c_{(i),j,k} = 0$$

for  $1 \leq k \leq D_{*,j}$ , and

$$c_{(i),j,k} = 0$$

for  $D_{*,j} < k \leq D_{*,j} + D'_{\text{ind},j}$ . But this forces  $\xi_j$  to vanish on  $V_{\text{ind},j}$ , a contradiction.  $\square$

We now apply the commutator argument to show that “independent” frequencies also ultimately have a trivial effect.

**Corollary 11.6** (Furstenberg-Weiss commutator argument, II). *Let  $w$  be an  $(r_* - 1)$ -fold iterated commutator of generators  $e_{j_1,k_1}, \dots, e_{j_{r_*},k_{r_*}}$  with  $1 \leq j_l \leq s-1$ ,  $1 \leq k_l \leq D_l$  for  $l = 1, \dots, r_*$  and  $j_1 + \dots + j_{r_*} = s-1$ . Suppose that at least one of the generators, say  $e_{j_1,k_1}$ , is an “independent” generator in the sense that  $D_{*,j_1} < k_1 \leq D_{*,j_1} + D'_{\text{ind},j_1}$ . Then  $(\phi_{(1)}(w), \text{id}, \text{id}, \text{id}, \text{id}) \in G_P$ .*

*Proof.* We may assume that  $k_l \leq D_{*,j_l}$  for all  $2 \leq l \leq r_*$ , as the claim would follow from Corollary 11.4 otherwise.

For  $e_{j_1,k_1}$ , we may invoke Lemma 11.5 and find an element  $g_{j_1,k_1}$  of  $G_P \cap G_{j_1}$  for which the first 4 coordinates are equal (modulo projection to  $\text{Horiz}_{j_1}(G_{(1)}) \times \text{Horiz}_{j_1}(G_{(2)}) \times \text{Horiz}_{j_1}(G_{(3)}) \times \text{Horiz}_{j_1}(G_{(4)})$ ) is equal to  $(\phi_{(1)}(e_{j_1,k_1}), \text{id}, \text{id}, \text{id})$ . For the other  $e_{j,k}$ , we can find elements  $g'_{j,k}$  of  $G_P \cap G_j$  for which the first coordinate is equal (modulo projection to  $\text{Horiz}_j(G_{(1)})$ ) to  $\phi_{(1)}(e_{j,k})$ . Taking commutators of  $g_{j_1,k_1}$  and  $g'_{j,k}$  in the order indicated by  $w$ , we obtain the claim.  $\square$

Combining Corollary 11.6 with Lemma 11.2 we obtain the second part of Theorem 11.1.

## 12. BUILDING A NILOBJECT

The aim of this section is to at last build an object coming from an  $s$ -step nilmanifold. Recall from the discussion in §7 that this object will be a multidegree  $(1, s-1)$ -nilcharacter  $\chi'(h, n)$ , and that this completes the proof of Theorem 7.2. This in turn was used iteratively to prove Theorem 7.1, the heart of our whole paper. It will then remain to supply the *symmetry argument*, which will take us from a 2-dimensional nilsequence to a 1-dimensional one; this will be accomplished in the next section.

Let  $f, H, \chi, (\chi_h)_{h \in H}$  be as in Theorem 7.2. If we apply Lemma 10.10, we obtain the following objects:

- A dense subset  $H'$  of  $H$ ;
- Dimension vectors  $\vec{D}_* = \vec{D}_{*,\text{ind}} + \vec{D}_{*,\text{rat}}$  and  $\vec{D}' = \vec{D}'_{\text{lin}} + \vec{D}'_{\text{ind}} + \vec{D}'_{\text{sml}}$ , which we write as  $\vec{D}_* = (D_{*,i})_{i=1}^{s-1}$ ,  $\vec{D}_{*,\text{ind}} = (D_{*,\text{ind},i})_{i=1}^{s-1}$ , etc.;
- A core horizontal frequency vector  $\mathcal{F}_* = (\xi_{*,i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D_{*,i}}$ , which is partitioned as  $\mathcal{F}_* = \mathcal{F}_{*,\text{ind}} \uplus \mathcal{F}_{*,\text{rat}}$ , with the indicated dimension vectors  $\vec{D}'_{\text{ind}}, \vec{D}'_{\text{rat}}$ ;
- A petal horizontal frequency vector  $\mathcal{F}'_h = (\xi'_{h,i,j})_{1 \leq i \leq s-1; 1 \leq j \leq D'_i}$ , which is partitioned as  $\mathcal{F}'_h = \mathcal{F}'_{h,\text{lin}} \uplus \mathcal{F}'_{h,\text{ind}} \uplus \mathcal{F}'_{h,\text{sml}}$ , which is a limit function of  $h$  and with the indicated dimension vectors  $\vec{D}'_{\text{lin}}, \vec{D}'_{\text{ind}}, \vec{D}'_{\text{sml}}$ ;
- Nilmanifolds  $G_h/\Gamma_h$  and  $G_{0,h}/\Gamma_{0,h}$  of degree-rank  $\leq (s-1, r_*)$  and  $\leq (s-1, r_*-1)$  respectively for each  $h \in H'$ , depending in a limit fashion on  $h$ ;
- Polynomial sequences  $g_h, g_{0,h} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G_h)_{\mathbb{N}})$  for each  $h \in H'$ , depending in a limit fashion on  $h$ ;
- Lipschitz functions  $F_h \in \text{Lip}^*(G_h/\Gamma_h \times G_{0,h}/\Gamma_{0,h} \rightarrow \overline{S^\omega})$  for each  $h \in H'$ , depending in a limit fashion on  $h$ ;
- a filtered  $\phi_h : G^{\vec{D}_* + \vec{D}'} \rightarrow G_h$  for each  $h \in H'$ , depending in a limit fashion on  $h$ ; and
- a character  $\eta_h : G_{(s-1, r_*)}^{\vec{D}_* + \vec{D}'} \rightarrow \mathbb{R}$  for each  $h \in H'$ , depending in a limit fashion on  $h$

that obey the following properties:

- For every  $1 \leq i \leq d$  and  $1 \leq j \leq D'_{i,\text{lin}}$ , there exists  $\alpha_{i,j} \in {}^*\mathbb{T}$  and  $\beta_{i,j} \in {}^*\mathbb{R}$  such that (10.2) holds, and furthermore that the map  $h \mapsto \xi'_{h,i,j}$  is a Freiman homomorphism on  $H'$ .
- For almost all additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ ,

$$\mathcal{F}_{*,\text{ind}} \uplus \biguplus_{i=1}^4 \mathcal{F}'_{h_i,\text{ind}} \uplus \biguplus_{i=1}^3 \mathcal{F}'_{h_i,\text{lin}}$$

is independent.

- We have the representation

$$\chi_h(n) = F_h(g_h(n)^* \Gamma_h, g_{0,h}(n)^* \Gamma_{0,h})$$

for every  $h \in H'$ .

- $\phi_h : G^{\vec{D}_* + \vec{D}'} \rightarrow G_h$  is a filtered homomorphism such that

$$F_h(\phi_h(t)x, x_0) = e(\eta_h(t))F_h(x, x_0) \quad (12.1)$$



- for all  $t \in G_{(s-1, r_*)}^{\bar{D}_* + \bar{D}'}$ ,  $x \in G_h/\Gamma_h$ , and  $x_0 \in G_{0,h}/\Gamma_{0,h}$ ;
- One has the Taylor coefficients

$$\text{Taylor}_i(g_h \Gamma_h) = \pi_{\text{Horiz}_i(G_h/\Gamma_h)}(\phi_h(\prod_{j=1}^{D_{*,i} + D'_i} e_{i,j}^{\xi_{h,i,j}})) \quad (12.2)$$

for all  $1 \leq i \leq s-1$ .

There are only countably many nilmanifolds  $G/\Gamma$  up to isomorphism, so by passing from  $H'$  to a dense subset using Lemma A.9 we may assume that

$$G_h/\Gamma_h = G/\Gamma \quad \text{and} \quad G_{0,h}/\Gamma_{0,h} = G_0/\Gamma_0$$

are independent of  $h$ . Similarly we may take  $\eta_h = \eta$  and  $\phi_h = \phi$  to be independent of  $h$ . From the Arzelà-Ascoli theorem, the space of possible  $F_h$  is totally bounded, and so (shrinking  $\varepsilon$  slightly if necessary) we may also assume that  $F_h = F$  is independent of  $h$ .

For  $j$  with  $1 \leq j \leq D_{*,i}$ , since  $\xi_{h,i,j}$  is independent of  $h$ , we can ensure that  $\xi_{h,i,j} = \gamma_{i,j}$  is also independent of  $h$ . Meanwhile, for  $D_{*,i} < j \leq D_{*,i} + D'_{i,\text{lin}}$ , from (10.2) we may assume that  $\xi_{h,i,j}$  takes the form

$$\xi_{h,i,j} = \{\alpha_{i,j}h\}\beta_{i,j} \bmod 1$$

for some  $\alpha_{i,j} \in {}^*\mathbb{T}$  and  $\beta_{i,j} \in {}^*\mathbb{R}$ . By passing to a dense subset of  $H'$  using the pigeonhole principle, we may assume for each  $i, j$ , that  $\{\alpha_{i,j}h\}$  is contained in a subinterval  ${}^*I_{i,j}$  around  ${}^*0$  of length at most  $1/10$  (say).

We now wish to apply Theorem 11.1 to obtain more convenient equivalent representatives (in  $\Xi_{\text{DR}}^{(s-1, r_*)}([N])$ )  $\tilde{\chi}_h$  for the nilcharacters  $\chi_h$ . Let  $\tilde{G}$  be the free Lie group generated by the generators  $\tilde{e}_{i,j}$  for  $1 \leq i \leq s-1$  and  $1 \leq j \leq D_{*,i} + D'_{i,\text{lin}}$  subject to the following relations:

- Any  $(r-1)$ -fold iterated commutator of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  with  $i_1 + \dots + i_r > s-1$  vanishes;
- Any  $(r-1)$ -fold iterated commutator of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  with  $i_1 + \dots + i_r = s-1$  and  $r > r_*$  vanishes;
- Any  $(r-1)$ -fold iterated commutator of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  in which  $j_l > D_{*,i_l}$  for at least two values of  $l$  vanishes.

We give this group a DR-filtration  $\tilde{G}_{\text{DR}}$  by defining  $\tilde{G}_{(d,r)}$  to be the group generated by the  $(r'-1)$ -fold iterated commutators of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_{r'}, j_{r'}}$  with  $i_1 + \dots + i_{r'} \geq d$  and  $r' \geq r$ . We then let  $\tilde{\Gamma}$  be the discrete group generated by the  $\tilde{e}_{i,j}$ ;  $\tilde{G}/\tilde{\Gamma}$  is then a nilmanifold of degree-rank  $\leq (s-1, r_*)$ .

Let  $G^*$  be the subgroup of  $G^{\bar{D}_* + \bar{D}'}$  generated by  $(r-1)$ -fold iterated commutators  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  with  $i_1 + \dots + i_r = s-1$  in which  $j_l > D_{*,i_l}$  for at least two values of  $l$ , or  $j_l > D_{*,i_l} + D'_{i_l, \text{lin}}$  for at least one value of  $l$ . Then  $G^*$  is a subgroup of the central group  $G_{(s-1, r_*)}^{\bar{D}_* + \bar{D}'}$  of  $G^{\bar{D}_* + \bar{D}'}$ , and  $\tilde{G}$  is isomorphic to the quotient of  $G^{\bar{D}_* + \bar{D}'}$  by  $G^*$ . We let  $\tilde{\phi} : G^{\bar{D}_* + \bar{D}'} \rightarrow \tilde{G}$  denote the quotient map. From Theorem 11.1, the character  $\eta : G_{(s-1, r_*)}^{\bar{D}_* + \bar{D}'} \rightarrow \mathbb{R}$  annihilates  $G^*$ , and thus descends to a vertical character  $\tilde{\eta} : \tilde{G}_{(s-1, r_*)} \rightarrow \mathbb{R}$ .

We select a function  $\tilde{F} \in \text{Lip}(\tilde{G}/\tilde{\Gamma} \rightarrow S^\omega)$  with vertical frequency  $\tilde{\eta}$ ; such a function can be built using the construction (6.3).

We then define the polynomial sequences  $g_0, \tilde{g}_h \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \tilde{G}_{\mathbb{N}})$  by the formulae

$$g_0(n) := \prod_{i=1}^{s-1} \prod_{j=1}^{D_{*,i}} \tilde{e}_{i,j}^{\gamma_{i,j}^{(n)}} \quad (12.3)$$

$$\tilde{g}_h(n) := \prod_{i=1}^{s-1} \prod_{j=D_{*,i}+1}^{D_{*,i}+D'_{\text{lin},i}} \tilde{e}_{i,j}^{\{\alpha_{i,j}h\}\beta_{i,j}^{(n)}} \quad (12.4)$$

and consider the nilcharacter

$$\tilde{\chi}_h(n) := \tilde{F}(g_0(n)\tilde{g}_h(n)^*\tilde{\Gamma}). \quad (12.5)$$

These nilcharacters are equivalent to  $\chi_h$  in  $\text{Symb}_{\text{DR}}^{(s-1,r_*)}([N])$ , as the following lemma shows.

**Lemma 12.1.** *For each  $h \in H'$ ,  $\chi_h$  and  $\tilde{\chi}_h$  are equivalent (as nilcharacters of degree-rank  $(s-1, r_*)$ ) on  $[N]$ .*

*Proof.* Fix  $h$ . It suffices to show that  $\chi_h \otimes \overline{\tilde{\chi}_h}$  is a nilsequence of degree  $< s-1$ . We can write this sequence as

$$n \mapsto F'_h(g'_h(n)^*\Gamma'), \quad (12.6)$$

where  $G' := G \times G_0 \times \tilde{G}$ ,  $\Gamma := \Gamma \times \Gamma_0 \times \tilde{\Gamma}$ ,  $g'_h \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G'_{\mathbb{N}})$  is the sequence

$$g'_h(n) := (g_h(n), g_{0,h}(n), g_0(n)\tilde{g}_h(n))$$

and  $F'_h \in \text{Lip}({}^*(G'/\Gamma'))$  is the function

$$F'_h(x, x_0, y) := F_h(x, x_0) \otimes \overline{\tilde{F}(y)}.$$

We define a DR-filtration  $G'_{\text{DR}}$  on  $G'$  by defining  $G'_{(d,r)}$  for  $(d,r) \in \text{DR}$  with  $r \geq 1$  to be the Lie group generated by the following sets:

- (i)  $G_{(d,r+1)} \times (G_0)_{(d,r)} \times \tilde{G}_{(d,r+1)}$ ;
- (ii)  $\{(\phi(g), \text{id}, \tilde{\phi}(g)) : g \in G_{(d,r)}^{\bar{D}_* + \bar{D}'}\}$ ,

with the convention that  $(d, d+1) = (d+1, 0)$ . We also set  $G'_{(d,0)} := G'_{(d,1)}$  for  $d \geq 1$ . One easily verifies that this is a filtration.

We claim that  $g'$  is polynomial with respect to this filtration. Indeed, the sequence  $n \mapsto (\text{id}, g_{0,h}(n), \text{id})$  is already polynomial in this filtration, so by Corollary B.4 it suffices to verify that the sequence

$$n \mapsto (g_h(n), \text{id}, g_0(n)\tilde{g}_h(n)) \quad (12.7)$$

is polynomial. We use Lemma B.9 to Taylor expand  $g_h(n) = \prod_{i=0}^{s-1} g_{h,i}^{(n)}$  where  $g_{h,i} \in G_{(i,0)}$ . From (12.2), one has

$$g_{h,i} = \phi\left(\prod_{j=1}^{D_{*,i}+D'_i} e_{i,j}^{\xi_{h,i,j}}\right) \bmod G_{(i,2)}.$$

By construction of the filtration of  $G'$ , this implies that

$$(g_{h,i}, \text{id}, \prod_{j=1}^{D_{*,i}+D'_i} e_{i,j}^{\xi_{h,i,j}} \bmod G^*) \in G'_{(i,1)}.$$

Applying Corollary B.4, we conclude that the sequence

$$n \mapsto (g_h(n), \text{id}, \prod_{i=0}^{s-1} (\prod_{j=1}^{D_{*,i}+D'_i} e_{i,j}^{\xi_{h,i,j}})^{\binom{n}{i}} \bmod G^*)$$

is polynomial with respect to the  $G'$  filtration. Applying the Baker-Campbell-Hausdorff formula repeatedly, and using (12.3), (12.4), we see that

$$n \mapsto \prod_{i=0}^{s-1} (\prod_{j=1}^{D_{*,i}+D'_i} e_{i,j}^{\xi_{h,i,j}})^{\binom{n}{i}} \bmod G^*$$

differs from the sequence  $n \mapsto g_0(n)\tilde{g}_h(n)$  by a sequence which is polynomial in the shifted filtration  $(\tilde{G}_{(d,r+1)})_{(d,r) \in \text{DR}}$ . We conclude that (12.7) is polynomial as required.

Next, we claim that  $F'_h$  is invariant with respect to the action of the central group

$$G'_{(s-1,r_*)} = \{(\phi(g), \text{id}, \tilde{\phi}(g)) : g \in G_{(s-1,r_*)}^{\tilde{D}}\}.$$

It suffices to check this for generators  $(\phi(w), \text{id}, w \bmod G^*)$ , where  $w$  is an  $(r_* - 1)$ -fold commutator of  $e_{i_1, j_1}, \dots, e_{i_{r_*}, j_{r_*}}$  in  $G^{\tilde{D}}$  with  $i_1 + \dots + i_{r_*} = s - 1$ . There are two cases. If one has  $j_l > D_{*,i_l} + D'_{\text{lin},i_l}$  for some  $l$ , then  $w$  lies in  $G^*$  and is also annihilated by  $\eta$ , and the claim follows from (12.1). If instead one has  $j_l \leq D_{*,i_l} + D'_{\text{lin},i_l}$  for all  $l$ , then the claim again follows from (12.1) together with the construction of  $\tilde{\eta}$  and  $\tilde{F}$ .

We may now quotient out  $G'_{(0,0)}$  by  $G'_{(s-1,r_*)}$  and obtain a representation of (12.6) as a nilsequence of degree-rank  $< (s - 1, r_*)$ , as desired.  $\square$

From this lemma and Lemma E.8(ii) we can express  $\chi_h$  as a bounded linear combination of  $\tilde{\chi}_h \otimes \psi_h$  for some nilsequence  $\psi_h$  of degree-rank  $\leq (s - 1, r_* - 1)$ . Thus, to prove Theorem 7.2 it suffices to show that there is a nilcharacter  $\tilde{\chi} \in \Xi^{(1,s-1)}(*\mathbb{Z}^2)$ , such that  $\tilde{\chi}_h(n) = \tilde{\chi}(h, n)$  for many  $h \in H'$  and all  $n \in [N]$ .

We illustrate the construction with an example. Let

$$G := G^{(2,0)} = \{e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}} : t_1, t_2, t_{12} \in \mathbb{R}\}$$

be the universal degree 2 nilpotent group (6.1) generated by  $e_1, e_2$ . Let  $F$  be the Lipschitz function in equation (6.3). Suppose

$$\chi_h(n) := F(g_h(n)^* \Gamma)$$

with  $g_h(n) := e_2^{\beta n} e_1^{\alpha_h n}$ , where  $\alpha_h := \{\delta h\}\gamma$ , and  $\alpha, \beta, \gamma \in {}^*\mathbb{R}$ . As computed in §6, we have

$$F_k(g_h(n)^* \Gamma) = \phi_k(\alpha_h n \bmod 1, \beta n \bmod 1) e(\{\alpha_h n\} \beta n)$$

for some Lipschitz function  $\phi_k : \mathbb{T}^2 \rightarrow \mathbb{C}$ . We would like to interpret the function  $(h, n) \mapsto \chi_h(n)$  as a nilcharacter in  $\Xi_{\text{Multi}}^{(1,2)}(*\mathbb{Z}^2)$ . The first task is to identify a subgroup  $G_{\text{petal}}$  of the group  $G$  representing that part of  $G$  that is “influenced by” the petal frequency  $\alpha_h$ ; more specifically, we take  $G_{\text{petal}}$  to be the subgroup of  $G$  generated by  $e_1$  and  $[e_1, e_2]$ , that is to say

$$G_{\text{petal}} = \langle e_1, [e_1, e_2] \rangle_{\mathbb{R}} = \{e_1^{t_1} [e_1, e_2]^{t_{12}} : t_1, t_{12} \in \mathbb{R}\}.$$

Note that  $G_{\text{petal}}$  is abelian and normal in  $G$ . In particular  $G$  acts on  $G_{\text{petal}}$  by conjugation, and we may form the semidirect product

$$G \ltimes G_{\text{petal}} := \{(g, g_1) : g \in G, g_1 \in G_{\text{petal}}\},$$

defining multiplication by

$$(g, g_1) \cdot (g', g'_1) = (gg', g_1^{g'} g'_1),$$

where  $a^b := b^{-1}ab$  denotes conjugation.

Now consider the action  $\rho$  of  $\mathbb{R}$  on  $G \ltimes G_{\text{petal}}$  defined by

$$\rho(t)(g, g_1) := (gg_1^t, g_1).$$

We may form a further semidirect product

$$G' := \mathbb{R} \ltimes_{\rho} (G \ltimes G_{\text{petal}}),$$

in which the product operation is defined by

$$(t, (g, g_1)) \cdot (t', (g', g'_1)) = (t + t', \rho(t')(g, g_1) \cdot (g', g'_1)).$$

$G'$  is a Lie group; indeed, one easily verifies that it is 3-step nilpotent. We give  $G'$  a  $\mathbb{N}^2$ -filtration:

$$\begin{aligned} G'_{(0,0)} &:= G' \\ G'_{(1,0)} &:= \{(t, (g, \text{id})) : t \in \mathbb{R}, g \in G_{\text{petal}}\} \\ G'_{(1,1)} &:= \{(0, (g, \text{id})) : g \in G_{\text{petal}}\}, \\ G'_{(1,2)} &:= \{(0, (g, \text{id})) : g \in [G, G]\}, \\ G'_{(0,1)} &:= \{(0, (g, g_1)) : g \in G_{\text{petal}}, g_1 \in G_{\text{petal}}\}, \\ G'_{(0,2)} &:= \{(0, (g, g_1)) : g, g_1 \in [G, G]\}, \end{aligned}$$

with  $G'_{i,j} := \{\text{id}\}$  for all other  $(i, j) \in \mathbb{N}^2$ . One easily verifies that this is a filtration. Inside  $G'$  we take the lattice

$$\Gamma' := \mathbb{Z} \ltimes_{\rho} (\Gamma \ltimes \Gamma_{\text{petal}}),$$

where  $\Gamma_{\text{petal}} := \Gamma \cap G_{\text{petal}}$ . Now consider the polynomial  $g' : \mathbb{Z}^2 \rightarrow G'$  defined by

$$g'(h, n) := (0, (e_2^{\beta n}, e_1^{\gamma n})) \cdot (\delta h, (\text{id}, \text{id}))$$

and observe that

$$\begin{aligned} g'(h, n)\Gamma' &= (0, (e_2^{\beta n}, e_1^{\gamma n})) \cdot (\{\delta h\}, (\text{id}, \text{id}))\Gamma' \\ &= (\{\delta h\}, (e_2^{\beta n} e_1^{\{\delta h\}\gamma n}, e_1^{\gamma n}))\Gamma'. \end{aligned}$$

For a dense subset  $H''$ ,  $\{\delta h\}$  is in a small interval  $I$ , and let  $\psi$  be a smooth cutoff function supported on  $2I$ . Take  $F' : G'/\Gamma' \rightarrow \mathbb{C}^D$  to be the function defined by

$$F'((t, (g, g'))\Gamma') := \psi(t)F(g\Gamma)$$

whenever  $t \in I$  and 0 otherwise. Then we have for  $h \in H''$

$$F'(g'(h, n)\tilde{\Gamma}) = F(e_2^{\beta n} e_1^{\{\delta h\}\gamma n} \Gamma) = \chi_h(n),$$

giving the desired representation of  $(h, n) \mapsto \chi_h(n)$  as an (almost) degree  $(1, 2)$  nilcharacter.

We now turn to the general case. Our construction shall proceed by an abstract algebraic construction. Let  $\tilde{G}_{\text{petal}}$  be the subgroup of  $\tilde{G}$  generated by  $(r - 1)$ -fold

( $r \geq 1$ ) iterated commutators of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  in which  $j_l > D_{*, i_l}$  for exactly one value of  $l$ . Then  $\tilde{G}_{\text{petal}}$  is a rational abelian normal subgroup of  $\tilde{G}$ . To see that  $\tilde{G}_{\text{petal}}$  is normal, one uses the equalities

$$\tilde{e}_{i,j}^{-1}[g, h]\tilde{e}_{i,j} = [\tilde{e}_{i,j}^{-1}g\tilde{e}_{i,j}, \tilde{e}_{i,j}^{-1}h\tilde{e}_{i,j}] \quad \text{and} \quad \tilde{e}_{i,j}^{-1}g\tilde{e}_{i,j} = g[g, \tilde{e}_{i,j}],$$

the commutator identities in equation (3.1), and the fact that any iterated commutators of  $\tilde{e}_{i_1, j_1}, \dots, \tilde{e}_{i_r, j_r}$  in which  $j_l > D_{*, i_l}$  for more than one value of  $l$  is trivial in  $\tilde{G}$ .

In particular,  $\tilde{G}$  acts on  $\tilde{G}_{\text{petal}}$  by conjugation, leading to the semidirect product  $\tilde{G} \ltimes \tilde{G}_{\text{petal}}$  of pairs  $(g, g_1)$  with the product

$$(g, g_1)(g', g'_1) := (gg', g_1^{g'} g'_1).$$

Next, let  $R$  be the commutative ring of tuples  $t = (t_{i,j})_{1 \leq i \leq s-1; D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}}$  with  $t_{i,j} \in \mathbb{R}$ , which we endow with the pointwise product. For each  $t \in R$ , we can define an homomorphism  $g \mapsto g^t$  on  $\tilde{G}$ , which we define on generators by mapping  $\tilde{e}_{i,j}$  to  $\tilde{e}_{i,j}^t$  for  $D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}$ , but preserving  $\tilde{e}_{i,j}$  for  $j \leq D_{*,i}$ . Such a homomorphism is well-defined as it preserves the defining relations of  $\tilde{G}$ . We observe the composition law

$$(g^t)^{t'} = g^{tt'}$$

for  $g \in \tilde{G}$  and  $t, t' \in R$ . Also, on the abelian subgroup  $\tilde{G}_{\text{petal}}$  on  $\tilde{G}$ , we see that

$$g^t g^{t'} = g^{t+t'} \quad (12.8)$$

as can be seen from the Baker-Campbell-Hausdorff formula (3.2). We can thus express

$$\tilde{g}_h(n) = g_1(n)^{\{\alpha h\}} \quad (12.9)$$

where  $g_1 \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (\tilde{G}_{\text{petal}})_{\mathbb{N}})$  is the polynomial sequence

$$g_1(n) := \prod_{i=1}^{s-1} \prod_{j=D_{*,i}+1}^{D_{*,i}+D'_{\text{lin},i}} \tilde{e}_{i,j}^{\beta_{i,j} \binom{n}{i}}$$

and  $\{\alpha h\} \in R$  is the element

$$\{\alpha h\} := (\{\alpha_{i,j} h\})_{1 \leq i \leq s-1; D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}}.$$

The homomorphism  $g \mapsto g^t$  preserves  $\tilde{G}_{\text{petal}}$ , and is the identity once  $\tilde{G}_{\text{petal}}$  is quotiented out. As a consequence we see that

$$(gg_1 g^{-1})^t = gg_1^t g^{-1} \quad (12.10)$$

for any  $g \in \tilde{G}$  and  $g_1 \in \tilde{G}_{\text{petal}}$ .

We can now define an action  $\rho$  of  $R$  (viewed now as an additive group) on  $\tilde{G} \ltimes \tilde{G}_{\text{petal}}$  by defining

$$\rho(t)(g, g_1) := (gg_1^t, g_1);$$

the properties (12.8), (12.10) ensure that this is indeed an action. We can then define the semi-direct product  $G' := R \ltimes_{\rho} (\tilde{G} \ltimes \tilde{G}_{\text{petal}})$  to be the set of pairs  $(t, (g, g_1))$  with the product

$$(t, (g, g_1))(t', (g', g'_1)) = (t + t', \rho(t')(g, g_1)(g', g'_1)).$$

This is a Lie group. We can give it a  $\mathbb{N}^2$ -filtration  $(G'_{(d_1, d_2)})_{(d_1, d_2) \in \mathbb{N}^2}$  as follows:

- (i) If  $d_1 > 1$ , then  $G'_{(d_1, d_2)} := \{\text{id}\}$ .

- (ii) If  $d_1 = 1$  and  $d_2 > 0$ , then  $G'_{(1,d_2)}$  consists of the elements  $(0, (g, \text{id}))$  with  $g \in \tilde{G}_{d_2} \cap \tilde{G}_{\text{petal}}$ .
- (iii) If  $d_1 = 1$  and  $d_2 = 0$ , then  $G'_{(1,0)}$  consists of the elements  $(t, (g, \text{id}))$  with  $t \in R$  and  $g \in \tilde{G}_{\text{petal}}$ .
- (iv) If  $d_1 = 0$  and  $d_2 > 0$ , then  $G'_{(0,d_2)}$  consists of the elements  $(0, (g, g_1))$  with  $g \in \tilde{G}_{d_2}$  and  $g_1 \in \tilde{G}_{\text{petal}} \cap \tilde{G}_{d_2}$ .
- (v)  $G'_{(0,0)} = G'$ .

One easily verifies that this is a filtration of degree  $\leq (1, s-1)$  with  $G'_{(0,0)} = G'$ .

We let  $\Gamma'$  be the subgroup of  $\tilde{G}$  consisting of pairs  $(t, (g, g_1))$  with  $g \in \tilde{\Gamma}$ ,  $g_1 \in \tilde{\Gamma}_{\text{petal}}$ , and with all coefficients of  $t$  integers. One easily verifies that  $\Gamma'$  is a cocompact subgroup of  $G'$ , and that the above  $\mathbb{N}^2$ -filtration of  $G'$  is rational with respect to  $\Gamma'$ , so that  $G'/\Gamma'$  has the structure of a filtered nilmanifold.

We consider the orbit  $\mathcal{O}' \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}^2}^2 \rightarrow (G'/\Gamma')_{\mathbb{N}^2})$  defined by

$$\mathcal{O}'(h, n) := (0, (g_0(n), g_1(n)))(\alpha h, (\text{id}, \text{id}))^* \Gamma',$$

where

$$\alpha h := (\alpha_{i,j} h)_{1 \leq i \leq s-1; D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}}.$$

As  $g_0, g_1$  were already known to be polynomial maps, and the linear map  $h \mapsto \alpha h$  is clearly polynomial also, we see from Corollary B.4 and the choice of filtration on  $G'$  that  $\mathcal{O}'$  is a polynomial orbit.

Now we simplify the orbit. Working on the abelian group  $R$ , we see that

$$(\alpha h, (\text{id}, \text{id}))^* \Gamma' = (\{\alpha h\}, (\text{id}, \text{id}))^* \Gamma',$$

and then commuting this with  $(0, (g_0(n), g_1(n)))$ , we obtain

$$\mathcal{O}'(h, n) = (\{\alpha h\}, (g_0(n)g_1(n)^{\{\alpha h\}}, g_1(n)))^* \Gamma'. \quad (12.11)$$

Recall that for many  $h \in H$  that each component  $\{\alpha_{i,j} h\}$  of  $\{\alpha h\}$  lies in an interval  $I_{i,j}$  of length at most  $1/10$ . Let  $2I_{i,j}$  be the interval of twice the length and with the same centre as  $I_{i,j}$ , and let  $\varphi_{i,j} : \mathbb{R} \rightarrow \mathbb{R}$  be a smooth cutoff function supported on  $I_{i,j}$ . We then define a function  $F' : G'/\Gamma' \rightarrow \mathbb{C}^\omega$  by setting

$$F'(((t_{i,j})_{1 \leq i \leq s-1; D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}}, (g, g_1))^* \Gamma') := \left( \prod_{i=1}^{s-1} \prod_{j=D_{*,i}+1}^{D_{*,i}+D'_{\text{lin},i}} \varphi_{i,j}(t_{i,j}) \right) \tilde{F}(g^* \tilde{\Gamma})$$

whenever  $(g, g_1) \in \tilde{G} \ltimes \tilde{G}_{\text{petal}}$  and  $t_{i,j} \in 2I_{i,j}$  for all  $1 \leq i \leq s-1$  and  $D_{*,i} < j \leq D_{*,i} + D'_{\text{lin},i}$ , with  $F'$  set equal to zero whenever no representation of the above form exists. One can easily verify that  $F'$  is well-defined and Lipschitz. Since  $\tilde{F}$  has vertical frequency  $\tilde{\eta}$ ,  $F'$  has vertical frequency  $\eta' : G'_{(1,s-1)} \rightarrow \mathbb{R}$ , defined by the formula

$$\eta'((0, (g, \text{id}))) := \tilde{\eta}(g)$$

for all  $g \in \tilde{G}_{s-1}$ . From (12.5), (12.9) and (12.11), we see that for many  $h \in H'$  we have

$$\tilde{\chi}_h(n) = F' \circ \mathcal{O}'(h, n)$$

for all  $n \in [N]$ . By construction,  $F' \circ \mathcal{O}' \in \Xi_{\text{Multi}}^{(1,s-1)}(*\mathbb{Z}^2)$ , and Theorem 7.1 follows.

## 13. THE SYMMETRY ARGUMENT

In this, the last section of the main part of the paper, we supply the symmetry argument, Theorem 7.4; we recall that statement now.

**Theorem 7.4.** *Let  $f \in L^\infty[N]$ , let  $H$  be a dense subset of  $[[N]]$ , and let  $\chi \in \Xi^{(1,s-1)}(*\mathbb{Z}^2)$  be such that  $\Delta_h f < (s-2)$ -correlates with  $\chi(h, \cdot)$  for all  $h \in H$ . Then there exists a nilcharacter  $\Theta \in \Xi^s(*\mathbb{Z})$  (with the degree filtration) and a nilsequence  $\Psi \in \text{Nil}^{<J}(*\mathbb{Z}^2)$  (with the multidegree filtration), with  $J$  given by the downset*

$$J := \{(i, j) \in \mathbb{N}^2 : i + j \leq s - 1\} \cup \{(i, s - i) : 2 \leq i \leq s\}, \quad (13.1)$$

such that  $\chi(h, n)$  is a bounded linear combination of  $\Theta(n + h) \otimes \overline{\Theta(n)} \otimes \Psi(h, n)$ .

**Example 13.1.** Suppose that  $s = 2$ ,  $\chi(h, n) = e(P(h, n))$ , and  $P(h, n) : *\mathbb{Z}^2 \rightarrow *\mathbb{R}$  is a symmetric bilinear form in  $n, h$ . Then observe that

$$\chi(h, n) = \Theta(n + h) \overline{\Theta(n)} \Psi(h, n) \quad (13.2)$$

where  $\Theta(n) := e(\frac{1}{2}P(n, n))$  and  $\Psi(h, n) := e(-\frac{1}{2}P(h, h))$ , which illustrates a special case of Theorem 7.4. More generally, if  $s \geq 2$  and  $\chi(h, n) = e(P(h, n, \dots, n))$  with  $P(h, n_1, \dots, n_{s-1}) : *\mathbb{Z}^s \rightarrow *\mathbb{R}$  a symmetric multilinear form, then we have (13.2) with  $\Theta(n) := e(\frac{1}{s}P(n, \dots, n))$ , and  $\Psi(h, n)$  a polynomial phase involving terms of multidegree  $(i, s - i)$  in  $h, n$  with  $2 \leq i \leq s$ . Thus we again obtain a special case of Theorem 7.4. Note how the symmetry of  $P$  is crucial in order to make these examples work, which explains why we refer to Theorem 7.4 as a symmetrisation result. Morally speaking, this type of symmetry property ultimately stems from the identity  $\Delta_h \Delta_k f = \Delta_k \Delta_h f$ . We remark that an analogous symmetrisation result was crucial to the analogous proof of GI(2) in [21] (see also [51]), although our arguments here are slightly different.

From the inclusions at the end of §6,  $\chi(h, n)$  is a nilcharacter on  $\mathbb{Z}^2$  (with the degree filtration) of degree  $\leq s$ . For similar reasons, any nilsequence  $\Psi(h, n)$  of degree  $\leq s - 1$  (using the degree filtration on  $\mathbb{Z}^2$ ) will automatically be of the form required for Theorem 7.4. In view of this and Lemma E.8, we see that it will suffice to obtain a factorisation of the form

$$[\chi]_{\Xi^s([N] \times [N])} = [\Theta(n + h)]_{\Xi^s([N] \times [N])} - [\Theta(n)]_{\Xi^s([N] \times [N])} + [\Psi(h, n)]_{\Xi^s([N] \times [N])}$$

where  $\Theta \in \Xi^s(*\mathbb{N})$  is a one-dimensional nilcharacter of degree  $\leq s$  (which automatically makes  $(h, n) \mapsto \Theta(n)$  and  $(h, n) \mapsto \Theta(n + h)$  two-dimensional nilcharacters of degree  $\leq s$ , by Lemma E.8(vi)), and  $\Psi \in \Xi^s(*\mathbb{N}^2)$  is a two-dimensional nilcharacter of multidegree

$$\subset \{(i, j) \in \mathbb{N}^2 : i + j \leq s; j \leq s - 2\}. \quad (13.3)$$

The set of classes  $[\Psi(h, n)]_{\Xi^s([N] \times [N])}$ , with  $\Psi$  of the above form, is a subgroup of the space  $\text{Symb}^s([N] \times [N])$  of all symbols of degree  $s$  nilcharacters on  $[N] \times [N]$ . Denoting the equivalence relation induced by these classes as  $\equiv$ , our task is thus to show that

$$[\chi]_{\Xi^s([N] \times [N])} \equiv [\Theta(n + h)]_{\Xi^s([N] \times [N])} - [\Theta(n)]_{\Xi^s([N] \times [N])}.$$

In view of Theorem E.10 and Lemma E.8 (vii), there is a nilcharacter  $\tilde{\chi}$  on  $*\mathbb{Z}^s$  of degree  $(1, \dots, 1)$  which is symmetric in the last  $s - 1$  variables, and such that

$$[\chi(h, n)]_{\Xi^s(*\mathbb{Z}^2)} = s[\tilde{\chi}(h, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)}. \quad (13.4)$$

Inspired by the polynomial identity

$$shn^{s-1} = (n+h)^s - n^s - \dots$$

where the terms in  $\dots$  are of degree  $s$  in  $h, n$  but of degree at most  $s-2$  in  $n$ , we now choose

$$\Theta(n) := \tilde{\chi}(n, \dots, n).$$

From Lemma E.8 (vi) we see that  $\Theta$  is a nilcharacter of degree  $\leq s$ . Our task is now to show that

$$\begin{aligned} [\tilde{\chi}(n+h, \dots, n+h)]_{\Xi^s([N] \times [N])} &- [\tilde{\chi}(n, \dots, n)]_{\Xi^s([N] \times [N])} - \\ &- s[\tilde{\chi}(h, n, \dots, n)]_{\Xi^s([N] \times [N])} \equiv 0. \end{aligned} \quad (13.5)$$

To manipulate this, we use the following lemma.

**Lemma 13.2** (Multilinearity). *Let  $\tilde{\chi}$  be a nilcharacter on  $\mathbb{Z}^s$  (with the multidegree filtration) of degree  $(1, \dots, 1)$ . Let  $m \geq 1$  be standard, and let  $L_1, \dots, L_s : \mathbb{Z}^m \rightarrow \mathbb{Z}$  and  $L'_1 : \mathbb{Z}^m \rightarrow \mathbb{Z}$  be homomorphisms. Then we have linearity in the first variable, in the sense that*

$$\begin{aligned} [\tilde{\chi}(L_1(\vec{n}) + L'_1(\vec{n}), L_2(\vec{n}), \dots, L_s(\vec{n}))]_{\Xi^s(*\mathbb{Z}^m)} &= [\tilde{\chi}(L_1(\vec{n}), L_2(\vec{n}), \dots, L_s(\vec{n}))]_{\Xi^s(*\mathbb{Z}^m)} \\ &+ [\tilde{\chi}(L'_1(\vec{n}), L_2(\vec{n}), \dots, L_s(\vec{n}))]_{\Xi^s(*\mathbb{Z}^m)}, \end{aligned}$$

where  $\vec{n} = (n_1, \dots, n_m)$  are the  $m$  independent variables of  $*\mathbb{Z}^m$ , and  $\mathbb{Z}^m$  is given the degree filtration. We similarly have linearity in the other  $s-1$  variables.

*Proof.* We prove the claim for the first variable, as the other cases follow from symmetry. From Lemma E.3 and Lemma E.8(vi), it will suffice to show that the expression

$$\tilde{\chi}(h_1 + h'_1, h_2, \dots, h_s) \otimes \overline{\tilde{\chi}}(h_1, h_2, \dots, h_s) \otimes \overline{\tilde{\chi}}(h'_1, h_2, \dots, h_s) \quad (13.6)$$

is a degree  $< s$  nilsequence in  $h_1, h'_1, h_2, \dots, h_s$  (using the degree filtration).

Write  $\tilde{\chi}(h_1, \dots, h_s) = F(g(h_1, \dots, h_s)*\Gamma)$ , where  $G/\Gamma$  is a  $\mathbb{N}^s$ -filtered nilmanifold of degree  $\leq (1, \dots, 1)$ ,  $F \in \text{Lip}(*G/\Gamma)$  has a vertical frequency, and  $g \in *\text{poly}(\mathbb{Z}_{\mathbb{N}^s}^s \rightarrow G_{\mathbb{N}^s})$ . Then the expression (13.6) takes the form

$$\tilde{F}(\tilde{g}(h_1, h'_1, h_2, \dots, h_s)*\Gamma^3)$$

where  $\tilde{g} : *\mathbb{Z}^{s+1} \rightarrow G^3$  is the map

$$\tilde{g}(h_1, h'_1, h_2, \dots, h_s) := (g(h_1 + h'_1, h_2, \dots, h_s), g(h_1, h_2, \dots, h_s), g(h'_1, h_2, \dots, h_s))$$

and  $\tilde{F} \in \text{Lip}(*G/\Gamma)^3$  is the map

$$\tilde{F}(x_1, x_2, x_3) = F(x_1) \otimes \overline{F(x_2)} \otimes \overline{F(x_3)}.$$

By Lemma B.9, we can expand

$$g(h_1, \dots, h_s) = \prod_{i_1, \dots, i_s \in \{0,1\}} g_{i_1, \dots, i_s}^{(h_1) \dots (h_s)}$$

for some  $g_{i_1, \dots, i_s} \in G_{(i_1, \dots, i_s)}$ , where we order  $\{0,1\}^s$  lexicographically (say).

We now  $\mathbb{N}$ -filter  $G^3$  by defining  $(G^3)_i$  to be the group generated by  $(G_{(i_1, \dots, i_s)})^3$  for all  $i_1, \dots, i_s \in \mathbb{N}$  with  $i_1 + \dots + i_s > i$ , together with the groups  $\{(g_1 g_2, g_1, g_2) : g_1, g_2 \in G_{(i_1, \dots, i_s)}\}$  for  $i_1 + \dots + i_s = i$ . From the Baker-Campbell-Hausdorff formula (3.2) one verifies that this is a rational filtration of  $G^3$ . From the Taylor expansion we also see that  $\tilde{g}$  is polynomial with respect to this filtration (giving  $\mathbb{Z}^{s+1}$  the



degree filtration). Finally, as  $F$  has a vertical character, we see that  $\tilde{F}$  is invariant with respect to the action of  $(G^3)_s = \{(g_1 g_2, g_1, g_2) : g_1, g_2 \in G_{(1, \dots, 1)}\}$ . Restricting  $G^3$  to  $(G^3)_0$  and quotienting out by  $(G^3)_s$  we obtain the claim.  $\square$

Using this lemma repeatedly, together with the symmetry of  $\tilde{\chi}$  in the final  $s-1$  variables, we see that we can expand

$$\begin{aligned} [\tilde{\chi}(n+h, \dots, n+h)]_{\Xi^s(*\mathbb{Z}^2)} &= \\ \sum_{j=0}^{s-1} \binom{s-1}{j} &([\tilde{\chi}(n, h, \dots, h, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)} + [\tilde{\chi}(h, h, \dots, h, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)}), \end{aligned}$$

where in the terms on the right-hand side, the final  $j$  coefficients are equal to  $n$ , the first coefficient is either  $n$  or  $h$ , and the remaining coefficients are  $h$ . Note that a term with  $j$   $h$  factors and  $(s-j)$   $n$  factors will have degree (13.3) and thus be negligible as long as  $j \geq 2$ . Neglecting these terms, we obtain the simpler expression

$$\begin{aligned} [\tilde{\chi}(n+h, \dots, n+h)]_{\Xi^s(*\mathbb{Z}^2)} &\equiv [\tilde{\chi}(n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)} + [\tilde{\chi}(h, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)} \\ &\quad + (s-1)[\tilde{\chi}(n, h, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^2)}. \end{aligned}$$

Comparing this with (13.3), we will be done as soon as we can show the symmetry property

$$(s-1)[\tilde{\chi}(h, n, \dots, n)]_{\Xi^s([N] \times [N])} = (s-1)[\tilde{\chi}(n, h, n, \dots, n)]_{\Xi^s([N] \times [N])}. \quad (13.7)$$

This property does not automatically follow from the construction of  $\tilde{\chi}$ . Instead, we must use the correlation properties of  $\chi$ , as follows.

By hypothesis and Lemma E.5, we have that for all  $h$  in a dense subset  $H$  of  $[N]$ , we can find a degree  $\leq s-2$  nilcharacter  $\varphi_h$  such that  $f_1(\cdot+h)f_2(\cdot)$  correlates with  $\chi(h, \cdot, \dots, \cdot) \otimes \varphi_h$ . By Corollary A.12, we may assume that the map  $h \mapsto \varphi_h$  is a limit map. We set  $\varphi_h = 0$  for  $h \notin H$ .

To use this information, we return<sup>7</sup> to Proposition 8.4. Invoking that proposition, we see that for many additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $[N]$ , the sequence

$$\begin{aligned} n \mapsto &\chi(h_1, n) \otimes \chi(h_2, n+h_1-h_4) \otimes \overline{\chi(h_3, n)} \otimes \overline{\chi(h_4, n+h_1-h_4)} \\ &\otimes \varphi_{h_1}(n) \otimes \varphi_{h_2}(n+h_1-h_4) \otimes \overline{\varphi_{h_3}(n)} \otimes \overline{\varphi_{h_4}(n+h_1-h_4)} \end{aligned}$$

is biased.

We make the change of variables  $(h_1, h_2, h_3, h_4) = (h+a, h+b, h+a+b, h)$  and then pigeonhole in  $h$ , to conclude the existence of an  $h_0$  for which

$$n \mapsto \tau(a, b, n) \otimes \varphi_{h_0+a}(n) \otimes \varphi_{h_0+b}(n+a) \otimes \overline{\varphi_{h_0+a+b}(n)} \otimes \overline{\varphi_{h_0}(n+a)}$$

is biased for many pairs  $a, b \in [2N]$ , where  $\tau = \tau_{h_0}$  is the expression

$$\tau(a, b, n) := \chi(h_0+a, n) \otimes \chi(h_0+b, n+a) \otimes \overline{\chi(h_0+a+b, n)} \otimes \overline{\chi(h_0, n+a)}. \quad (13.8)$$

Henceforth  $h_0$  is fixed, and we will suppress the dependence of various functions on this parameter. From Lemma E.3,  $\tau$  is a degree  $\leq 3$  nilcharacter on  ${}^*\mathbb{Z}^3$  (with the degree filtration). We record its top order symbol:

<sup>7</sup>Here is a key place where we use the hypothesis  $s \geq 3$  (the other is Lemma 10.9). For  $s = 2$  the lower order terms in Proposition 8.4 are useless; however a variant of the argument below still works, see [21].

**Lemma 13.3.** *We have*

$$[\tau(a, b, n)]_{\Xi^s(*\mathbb{Z}^3)} \equiv s(s-1)[\tilde{\chi}(b, a, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^3)}$$

where  $by \equiv$  we are quotienting by all symbols of degree  $\leq s-3$  in  $n$ .

*Proof.* From (13.4), (13.8), Lemma E.3 and Lemma E.8 one has

$$\begin{aligned} [\tau(a, b, n)]_{\Xi^s(*\mathbb{Z}^3)} &= s([\tilde{\chi}(a, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^3)} + [\tilde{\chi}(b, n+a, \dots, n+a)]_{\Xi^s(*\mathbb{Z}^3)} - \\ &\quad - [\tilde{\chi}(a+b, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^3)}). \end{aligned}$$

Applying Lemma 13.2 in the first variable we simplify this as

$$s([\tilde{\chi}(b, n+a, \dots, n+a)]_{\Xi^s(*\mathbb{Z}^3)} - [\tilde{\chi}(a, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^3)}).$$

Applying Lemma 13.2 in all the other variables and gathering terms using the symmetry of  $\tilde{\chi}$  in those variables, we arrive at

$$\sum_{j=0}^{s-2} s \binom{s-1}{j} [\tilde{\chi}(b, a, \dots, a, n, \dots, n)]_{\Xi^s(*\mathbb{Z}^3)},$$

where there are  $j$  occurrences of  $n$  and  $s-1-j$  occurrences of  $a$ . All the terms with  $j < s-2$  are of degree  $\leq s-2$  in  $n$ , and the claim follows.  $\square$

From Lemma E.8, we know that  $\varphi_{h_0+b}(n+a)$  is a bounded linear combination of  $\varphi_{h_0+b}(n) \otimes \psi_{a,b}(n)$  for some degree  $\leq s-3$  nilsequence  $\psi_{a,b}$ . Similarly for  $\varphi_{h_0}(n+a)$ . We conclude that

$$n \mapsto \tau(a, b, n) \otimes \varphi_{h_0+a}(n) \otimes \varphi_{h_0+b}(n) \otimes \overline{\varphi_{h_0+a+b}(n)} \otimes \overline{\varphi_{h_0}(n)}$$

is  $\leq (s-3)$ -biased for many  $a, b \in [[2N]]$ .

We will now eliminate the  $\varphi_h$  terms in order to focus attention on  $\tau$ . Applying Corollary A.12, we may thus find a scalar degree  $\leq s-3$  nilsequence  $\psi_{a,b}$  depending in a limit fashion on  $a, b \in [[2N]]$ , such that

$$\begin{aligned} |\mathbb{E}_{a,b \in [[2N]]; n \in [N]} \tau(a, b, n) \otimes \varphi_{h_0+a}(n) \otimes \varphi_{h_0+b}(n) \otimes \overline{\varphi_{h_0+a+b}(n)} \otimes \\ \otimes \overline{\varphi_{h_0,k'}(n+a)} \psi_{a,b}(n)| \gg 1. \end{aligned}$$

We pull out the  $b$ -independent factors  $\varphi_{h_0+a}(n) \otimes \overline{\varphi_{h_0}(n)}$  and Cauchy-Schwarz in  $a, n$  to conclude that

$$\begin{aligned} |\mathbb{E}_{a,b,b' \in [[2N]]; n \in [N]} \tau(a, b, n) \otimes \overline{\tau(a, b', n)} \otimes \varphi_{h_0+b}(n) \otimes \overline{\varphi_{h_0+b'}(n)} \\ \otimes \overline{\varphi_{h_0+a+b}(n)} \otimes \varphi_{h_0+a+b'}(n) \psi_{a,b,b'}(n)| \gg 1, \end{aligned}$$

where  $(a, b, b') \mapsto \psi_{a,b,b'}$  is a limit map assigning a scalar degree  $\leq s-3$  nilsequence to each  $a, b, b'$ . Next, we make the substitution  $c := a + b + b'$  and conclude that

$$\begin{aligned} |\mathbb{E}_{c,b,b' \in [[3N]]; n \in [N]} \tau(c-b-b', b, n) \otimes \overline{\tau(c-b-b', b', n)} \\ \otimes \varphi_{h_0+b}(n) \otimes \overline{\varphi_{h_0+b'}(n)} \otimes \overline{\varphi_{h_0+c-b'}(n)} \varphi_{h_0+c-b}(n) \psi'_{c,b,b'}(n)| \gg 1 \end{aligned}$$

where  $(c, b, b') \mapsto \psi'_{c,b,b'}$  is a limit map assigning a scalar degree  $\leq s-3$  nilsequence to each  $c, b, b'$ . By the pigeonhole principle, we can thus find a  $c_0$  such that

$$|\mathbb{E}_{b,b' \in [[3N]]; n \in [N]} \alpha(b, b', n) \otimes \varphi'_b(n) \otimes \overline{\varphi'_{b'}(n)} \psi'_{c_0,b,b'}(n)| \gg 1 \quad (13.9)$$

where  $\alpha = \alpha_{c_0}$  is the form

$$\alpha(b, b', n) := \tau(c_0 - b - b', b, n) \otimes \overline{\tau(c_0 - b - b', b', n)} \quad (13.10)$$

and  $\varphi'_b = \varphi'_{b,c_0}$  is the quantity

$$\varphi'_b(n) := \varphi_{h_0+b,k}(n) \otimes \overline{\varphi_{h_0+c_0-b}(n)}.$$

We fix this  $c_0$ . Again by Lemma E.3,  $\alpha$  is a degree  $\leq s$  nilcharacter on  ${}^*\mathbb{Z}^3$ , and we pause to record its symbol in the following lemma.

**Lemma 13.4.** *We have*

$$[\alpha(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)} \equiv -s(s-1)[\tilde{\chi}(b+b', b-b', n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)}$$

where  $by \equiv$  we are quotienting by all symbols of degree  $\leq s-3$  in  $n$ .

*Proof.* From (13.10) and Lemma E.8 we can write the left-hand side as

$$[\tau(-b-b', b, n)]_{\Xi^s({}^*\mathbb{Z}^3)} - [\tau(-b-b', b', n)]_{\Xi^s({}^*\mathbb{Z}^3)}.$$

Applying (13.3), we can write this as

$$s(s-1)([\tilde{\chi}(-b-b', b, n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)} - [\tilde{\chi}(-b-b', b', n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)}).$$

The claim then follows from some applications of Lemma 13.2.  $\square$

We return now to (13.9), and Cauchy-Schwarz in  $b', n$  to eliminate the  $\varphi'_{b'}(n)$  factor, yielding

$$|\mathbb{E}_{b_1, b_2, b' \in [[3N]]; n \in [N]} \alpha(b_1, b', n) \otimes \overline{\alpha(b_2, b', n)} \otimes \varphi'_{b_1}(n) \otimes \overline{\varphi'_{b_2}(n)} \psi''_{b_1, b_2, b'}(n)| \gg 1$$

where  $(b_1, b_2, b') \mapsto \psi''_{b_1, b_2, b'}$  is a limit map assigning a scalar degree  $\leq s-3$  nilsequence to each  $b_1, b_2, b'$ . Finally, we Cauchy-Schwarz in  $b_1, b_2, n$  to eliminate the  $\varphi'_{b_1}(n) \overline{\varphi'_{b_2}(n)}$  factor, yielding

$$|\mathbb{E}_{b_1, b_2, b'_1, b'_2 \in [[3N]]; n \in [N]} \alpha(b_1, b'_1, n) \otimes \overline{\alpha(b_2, b'_1, n)} \otimes \overline{\alpha(b_1, b'_2, n)} \otimes \alpha(b_2, b'_2, n) \psi''_{b_1, b_2, b'_1, b'_2}(n)| \gg 1.$$

Note how the  $\varphi$  terms have now been completely eliminated. To eliminate the  $\psi''$  terms, we first use the pigeonhole principle to find  $b_0, b'_0$  such that

$$|\mathbb{E}_{b, b' \in [[3N]]; n \in [N]} \alpha'(b, b', n) \psi''_{b, b_0, b', b'_0}(n)| \gg 1 \quad (13.11)$$

where  $\alpha' = \alpha'_{b_0, b'_0}$  is the expression

$$\alpha'(b, b', n) := \alpha(b, b', n) \otimes \overline{\alpha(b_0, b', n)} \otimes \overline{\alpha(b, b'_0, n)} \otimes \alpha(b_0, b'_0, n). \quad (13.12)$$

We fix this  $b_0, b'_0$ . Again,  $\alpha'$  is a degree  $\leq s$  nilcharacter on  ${}^*\mathbb{Z}^3$ . From Lemma 13.4 and Lemma 13.2 (and using Lemma E.8 to eliminate shifts by  $b_0$ ) we conclude

$$[\alpha'(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)} \equiv s(s-1)([\tilde{\chi}(b, b', n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)} - [\tilde{\chi}(b', b, n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)}). \quad (13.13)$$

Note the similarity here with (13.7).

From (13.11), we conclude that the sequence  $n \mapsto \alpha'(b, b', n)$  is  $\leq s-3$ -biased for many  $b, b' \in [[3N]]$ . Applying Proposition 5.6, we conclude that

$$\|\alpha'(b, b', n)\|_{U^{s-2}[N]} \gg 1$$

for many  $b, b' \in [[3N]]$ . We conclude (using Corollary A.6 to obtain the needed uniformity) that

$$\mathbb{E}_{b, b' \in [[3N]]} \|\alpha'(b, b', n)\|_{U^{s-2}[N]}^{2^{s-2}} \gg 1.$$

By definition of the Gowers norm, this implies that

$$|\mathbb{E}_{b, b', h_1, \dots, h_{s-2} \in [[3N]]; n \in [N]} \sigma(b, b', h_1, \dots, h_{s-2}, n) 1_{\Omega}(h_1, \dots, h_{s-2}, n)| \gg 1, \quad (13.14)$$

where  $\Omega$  is the polytope

$$\Omega := \{(h_1, \dots, h_{s-2}, n) : n + \sum_{j=1}^{s-2} \omega_j h_{s-2} \in [N] \text{ for all } \omega \in \{0, 1\}^{s-2}\}$$

and  $\sigma$  is the expression

$$\sigma(b, b', h_1, \dots, h_{s-2}, n) := \bigotimes_{\omega \in \{0, 1\}^{s-2}} \mathcal{C}^{|\omega|} \alpha'(b, b', n + \sum_{j=1}^{s-2} \omega_j h_{s-2}), \quad (13.15)$$

with  $\mathcal{C}$  being the conjugation map.

From Lemma E.3,  $\sigma$  is a nilcharacter of degree  $s$  on  ${}^*\mathbb{Z}^{s+1}$ . In the following lemma we compute its symbol.

**Lemma 13.5.** *We have*

$$\begin{aligned} [\sigma(b, b', h_1, \dots, h_{s-2}, n)]_{\Xi^s({}^*\mathbb{Z}^{s+1})} &= s!([\tilde{\chi}(b, b', h_1, \dots, h_{s-2})]_{\Xi^s({}^*\mathbb{Z}^{s+1})} \\ &\quad - [\tilde{\chi}(b', b, h_1, \dots, h_{s-2})]_{\Xi^s({}^*\mathbb{Z}^{s+1})}). \end{aligned} \quad (13.16)$$

*Proof.* From (13.15) and Lemma E.8 we can write the left-hand side as

$$\sum_{\omega \in \{0, 1\}^{s-2}} (-1)^{|\omega|} [\alpha'(b, b', n + \sum_{j=1}^{s-2} \omega_j h_{s-2})]_{\Xi^s({}^*\mathbb{Z}^{s+1})}; \quad (13.17)$$

one should think of this as an  $s - 2$ -fold “derivative” of  $[\alpha'(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)}$  in the  $n$  variable.

From (13.13) we can write

$$\begin{aligned} [\alpha'(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)} &= s(s-1)([\tilde{\chi}(b, b', n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)} - [\tilde{\chi}(b', b, n, \dots, n)]_{\Xi^s({}^*\mathbb{Z}^3)}) \\ &\quad + [\beta(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)} \end{aligned}$$

where  $\beta$  is of degree at most  $s - 3$  in  $n$ . In fact, by inspection of the derivation of  $\beta$ , and heavy use of Lemma 13.2, one can express  $[\beta(b, b', n)]_{\Xi^s({}^*\mathbb{Z}^3)}$  as a linear combination of classes of the form

$$[\tilde{\chi}(n_1, \dots, n_s)]_{\Xi^s({}^*\mathbb{Z}^3)}$$

where each of  $n_1, \dots, n_s$  is equal to either  $b, b'$ , or  $n$ , with at most  $s - 3$  copies of  $n$  occurring. If one then substitutes this expansion into (13.17) and applies Lemma 13.2 repeatedly, one obtains the claim.  $\square$

On the other hand, from (13.14) and Lemma E.11, we see that on  $[[3N]]^{s+1}$ ,  $\sigma$  is equal to a nilsequence of degree  $\leq s - 1$ , and thus by Lemma E.8

$$[\sigma(b, b', h_1, \dots, h_{s-2}, n)]_{\Xi^s([3N]^{s+1})} = 0$$

and thus by Lemma (13.16)

$$s!([\tilde{\chi}(b, b', h_1, \dots, h_{s-2})]_{\Xi^s([3N]^{s+1})} - [\tilde{\chi}(b', b, h_1, \dots, h_{s-2})]_{\Xi^s([3N]^{s+1})}) = 0.$$

Applying Lemma E.3 we conclude that

$$s!([\tilde{\chi}(h, n, \dots, n)]_{\Xi^s([N] \times [N])} - [\tilde{\chi}(n, h, n, \dots, n)]_{\Xi^s([N] \times [N])}) = 0.$$

The claim (13.7) now follows from Lemma E.13. The proof of Theorem 7.4 is now complete.

## APPENDIX A. BASIC THEORY OF ULTRALIMITS

In this appendix we review the machinery of ultralimits.

We will assume the existence of a *standard universe*  $\mathfrak{U}$  which contains all the objects and spaces of interest for Theorem 1.3, such as real numbers, subsets of real numbers, functions from  $[N]$  to  $\mathbb{C}$  for finite  $N \in \mathbb{N}$ , nilmanifolds (or more precisely, a representative from each equivalence class of nilmanifolds), and so forth. The precise construction of this universe is not important, so long as it forms a set. We refer to objects and spaces inside the standard universe as *standard objects* and *standard spaces*, with the latter being sets whose elements are in the former category. Thus for instance, elements of  $\mathbb{N}$  are standard natural numbers, the Heisenberg nilmanifold  $\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix} / \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$  is a standard nilmanifold (consisting entirely of standard points), and so forth.

The one technical ingredient we need is the following:

**Lemma A.1** (Ultrafilter lemma). *There exists a collection  $p$  of subsets of the natural numbers  $\mathbb{N}$  with the following properties:*

- (i) (Monotonicity) *If  $A \in p$  and  $B \supset A$ , then  $B \in p$ .*
- (ii) (Closure under intersection) *If  $A, B \in p$ , then  $A \cap B \in p$ .*
- (iii) (Maximality) *If  $A \subset \mathbb{N}$ , then either  $A \in p$  or  $\mathbb{N} \setminus A \in p$ , but not both.*
- (iv) (Non-principality) *If  $A \in p$ , and  $A'$  is formed from  $A$  by adding or deleting finitely many elements to or from  $A$ , then  $A' \in p$ .*

*Proof.* The collection of subsets of  $\mathbb{N}$  which are cofinite (i.e. whose complement is finite) already obeys the monotonicity, closure under intersection, and non-principality properties. Using Zorn's lemma<sup>8</sup>, one can enlarge this collection to a maximal collection, which then obeys all the required properties.  $\square$

Throughout the paper, we fix a non-principal ultrafilter  $p$ . A property  $P(\mathbf{n})$  depending on a natural number  $\mathbf{n}$  is said to hold *for  $\mathbf{n}$  sufficiently close to  $p$*  if the set of  $\mathbf{n}$  for which  $P(\mathbf{n})$  holds lies in  $p$ .

Once we have fixed this ultrafilter we can define limit objects and spaces as follows.

**Definition A.2** (Limit objects). Given a sequence  $(x_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$  of standard objects in  $\mathfrak{U}$ , we define their *ultralimit*  $\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}}$  to be the equivalence class of all sequences  $(y_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$  of standard objects in  $\mathfrak{U}$  such that  $x_{\mathbf{n}} = y_{\mathbf{n}}$  for  $\mathbf{n}$  sufficiently close to  $p$ . Note that the ultralimit  $\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}}$  can also be defined even if  $x_{\mathbf{n}}$  is only defined for  $\mathbf{n}$  sufficiently close to  $p$ .

An ultralimit of standard natural numbers is known as a *limit natural number*, an ultralimit of standard real numbers is known as a *limit real number*, etc.

For any standard object  $x$ , we identify  $x$  with its own ultralimit  $\lim_{\mathbf{n} \rightarrow p} x$ . Thus, every standard natural number is a limit natural number, etc.

---

<sup>8</sup>By using this lemma, our results thus rely on the axiom of choice, which we will of course assume throughout this paper. On the other hand, it is tedious but straightforward to rephrase the inverse conjecture (Conjecture 1.2) in the language of Peano arithmetic (e.g. using Mal'cev bases [48] to represent a nilmanifold, and approximating a Lipschitz function by a piecewise linear one). Applying a famous theorem of Gödel [15], we then conclude that Conjecture 1.2 is provable in ZFC if and only if it is provable in ZF. In fact, it is possible (with some effort) to directly translate these ultrafilter arguments to a (lengthier) argument in which ultrafilters or the axiom of choice is not used. We will not do so here, though, as the translation is quite tedious.

Any operation or relation on standard objects can be extended to limit objects in the obvious manner. For instance, the sum of two limit real numbers  $\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}}$ ,  $\lim_{\mathbf{n} \rightarrow p} y_{\mathbf{n}}$  is the limit real number

$$\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}} + \lim_{\mathbf{n} \rightarrow p} y_{\mathbf{n}} = \lim_{\mathbf{n} \rightarrow p} (x_{\mathbf{n}} + y_{\mathbf{n}}),$$

and the statement  $\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}} < \lim_{\mathbf{n} \rightarrow p} y_{\mathbf{n}}$  means that  $x_{\mathbf{n}} < y_{\mathbf{n}}$  for all  $\mathbf{n}$  sufficiently close to  $p$ .

A famous theorem of Łoś asserts that any statement in first-order logic which is true for standard objects is automatically true for limit objects as well. For instance, the standard real numbers form an ordered field, and so the limit real numbers do also, because the axioms of an ordered field can be phrased in first-order logic. We will use this theorem in the sequel without further comment.

**Definition A.3** (Limit spaces and functions). Let  $(X_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$  be a sequence of standard spaces  $X_{\mathbf{n}}$  in  $\mathfrak{U}$  indexed by the natural numbers. The *ultrapower*  $\prod_{\mathbf{n} \rightarrow p} X_{\mathbf{n}}$  of the  $X_{\mathbf{n}}$  is defined to be the space of all ultralimits  $\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}}$ , where  $x_{\mathbf{n}} \in X_{\mathbf{n}}$  for all  $\mathbf{n}$ . Note  $X_{\mathbf{n}}$  only needs to be well-defined for  $\mathbf{n}$  sufficiently close to  $p$  in order for the ultraproduct to be well-defined. If  $X$  is a set, the set  $\prod_{\mathbf{n} \rightarrow p} X$  is known as the *ultrapower* of  $X$  and is denoted  $*X$ . Thus for instance  $*\mathbb{N}$  is the space of all limit natural numbers,  $*\mathbb{R}$  is the space of all limit reals, etc.

We define a *limit set* to be an ultraproduct of sets, a *limit group* to be an ultraproduct of groups, a *limit finite set* to be an ultraproduct of finite sets, and so forth. A *limit subset* of a limit set  $X = \prod_{\mathbf{n} \rightarrow p} X_{\mathbf{n}}$  is a limit set of the form  $Y = \prod_{\mathbf{n} \rightarrow p} Y_{\mathbf{n}}$ , where  $Y_{\mathbf{n}}$  is a standard subset of  $X_{\mathbf{n}}$  for all  $\mathbf{n}$  sufficiently close to  $p$ .

Given a sequence of standard functions  $f_{\mathbf{n}} : X_{\mathbf{n}} \rightarrow Y_{\mathbf{n}}$  between standard sets  $X_{\mathbf{n}}, Y_{\mathbf{n}}$ , we can form the *ultralimit*  $f = \lim_{\mathbf{n} \rightarrow p} f_{\mathbf{n}}$  to be the function  $f : \prod_{\mathbf{n} \rightarrow p} X_{\mathbf{n}} \rightarrow \prod_{\mathbf{n} \rightarrow p} Y_{\mathbf{n}}$  defined by the formula

$$f(\lim_{\mathbf{n} \rightarrow p} x_{\mathbf{n}}) := \lim_{\mathbf{n} \rightarrow p} f_{\mathbf{n}}(x_{\mathbf{n}}).$$

We refer to  $f$  as a *limit function* or *limit map*, and say that  $f(x)$  depends *in a limit fashion* on  $x$ .

*Remark.* In the nonstandard analysis literature, limit natural numbers are known as *nonstandard natural numbers*, limit sets are known as *internal sets*, and limit functions are known as *internal functions*. We have chosen the limit terminology instead as we believe that it is less confusing and emphasises the role of ultralimits in the subject.

It is important to note that not every subset of a limit set is again a limit set, for instance  $\mathbb{N}$  is not a limit subset of  $*\mathbb{N}$  (this fact is known as the *overspill principle*). Indeed, one can think of the limit subsets of a limit set as being analogous to the measurable subsets of a measure space. In a similar vein, not every function between two limit sets is a limit function; in this regard, limit functions are analogous to measurable functions.

*Example.* (Pigeonhole principle) If  $X$  is finite, then  $*X = X$ . This is ultimately because if the natural numbers is partitioned into finitely many classes, then exactly one of those classes lies in  $p$ . In particular, we see that every standard finite set is a limit finite set. However, the converse is not true. For instance, if  $N$  is the limit

natural number  $N := \lim_{\mathbf{n} \rightarrow p} \mathbf{n}$ , then the limit set

$$[N] := \{n \in {}^*\mathbb{N} : 1 \leq n \leq N\} = \prod_{\mathbf{n} \rightarrow p} [\mathbf{n}]$$

is a limit finite set, but not a finite set.

*Example.* One has the identifications  ${}^*\mathbb{T} = ({}^*\mathbb{R})/({}^*\mathbb{Z})$  and  ${}^*(\mathbb{R}^k) = ({}^*\mathbb{R})^k$  for any standard  $k$ , so one can talk about the limit unit circle  ${}^*\mathbb{T}$  or the limit vector space  ${}^*\mathbb{R}^k$  without ambiguity. We will refer to elements of  ${}^*\mathbb{T}$  as *frequencies*.

*Example.* Every standard function  $f : X \rightarrow Y$  can be identified with its ultra-limit  $f : {}^*X \rightarrow {}^*Y$ , thus for instance the fundamental character  $e$  is a limit function from  ${}^*\mathbb{R}$  (or  ${}^*\mathbb{T}$ ) to  ${}^*\mathbb{C}$ , and the fractional part function  $\{\}$  is a limit function from  ${}^*\mathbb{R}$  to  ${}^*I_0$ .

*Remark.* A limit finite set  $A = \lim_{\mathbf{n} \rightarrow p} A_{\mathbf{n}}$  has an *limit cardinality*  $|A|$ , defined by the formula

$$|A| := \lim_{\mathbf{n} \rightarrow p} |A_{\mathbf{n}}|.$$

Of course,  $|A|$  is a limit natural number, and not a natural number in general. Thus for instance, if  $N$  is a limit natural number, then the limit finite set  $[N]$  has a limit cardinality of  $N$  (despite being uncountable in the standard sense).

ASYMPTOTIC NOTATION. By taking ultralimits, one can formalise asymptotic notation, such as the  $O()$  notation, in a manner that requires no additional quantifiers.

**Definition A.4** (Asymptotic notation). A limit complex number  $X$  is said to be *bounded* if one has  $|X| \leq C$  for some standard real number  $C$ , in which case we also write  $X = O(1)$  or  $|X| \ll 1$ . More generally, given a limit complex number  $X$  and limit non-negative number  $Y$ , we write  $|X| \ll Y$ ,  $Y \gg |X|$ , or  $X = O(Y)$  if one has  $|X| \leq CY$  for some standard real number  $C$ . We write  $X = o(Y)$  if one has  $|X| \leq \varepsilon Y$  for every standard  $\varepsilon > 0$ . Observe that for any  $X, Y$  with  $Y$  positive, one has either  $|X| \gg Y$  or  $X = o(Y)$ . We say that  $X$  is *infinitesimal* if  $X = o(1)$ , and *unbounded* if  $1/X = o(1)$ . Thus for instance any limit complex number  $X$  will either be bounded or unbounded.

In a similar spirit, if  $x \in {}^*V$  is a limit element of a standard topological space  $V$ , we say that  $x$  is *bounded* if  $x$  is a limit element of standard compact subset  $K$  of  $V$  (i.e.  $x \in {}^*K$ ), and *unbounded* otherwise. The set of all bounded elements of  ${}^*V$  will be denoted  $\overline{V}$ .

*Example.* The limit real  $\lim_{\mathbf{n} \rightarrow p} 1/\mathbf{n}$  defines an infinitesimal, but non-zero, limit real number  $x$ ; its reciprocal  $\lim_{\mathbf{n} \rightarrow p} \mathbf{n}$  is an unbounded limit real.

*Example.* Any bounded element of a discrete standard space is standard, by our example on the pigeonhole principle. In particular, bounded integers are automatically standard:  $\overline{\mathbb{Z}} = \mathbb{Z}$ . On the other hand, bounded elements in a continuous space need not be standard, as the example  $\lim_{\mathbf{n} \rightarrow p} 1/\mathbf{n}$  shows.

From the Bolzano-Weierstrass theorem, every bounded limit real number can be expressed uniquely as the sum of a standard real number and an infinitesimal, which may help explain the notation  $\overline{\mathbb{R}}$ . Note that  $\overline{\mathbb{R}}$  contains the limit fundamental domain  ${}^*I_0$ . Similarly,  $\overline{\mathbb{C}}$  contains the limit unit circle  ${}^*S^1 = \overline{S^1} = \{z \in \overline{\mathbb{C}} : |z| = 1\}$ , where  $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ .

*Example.* For any standard  $D \in \mathbb{N}^+$ , we endow  $\mathbb{C}^D$  with the Euclidean norm

$$|(z_1, \dots, z_D)| := (|z_1|^2 + \dots + |z_D|^2)^{1/2}.$$

Then we have  $\overline{\mathbb{C}^D} = \overline{\mathbb{C}}^D$ : an element  $(z_1, \dots, z_D) \in {}^*\mathbb{C}^D$  is bounded if and only if each component is bounded.

One modest advantage of the ultralimit framework is that one can rigorously work with such equivalence relations as “ $x$  and  $y$  differ by  $O(1)$ ”, for instance by quotienting  ${}^*\mathbb{R}$  by the subring  $\overline{\mathbb{R}}$ ; in the finitary setting, this relation is only “morally” an equivalence relation (because of the need to quantify the constants in the  $O()$  notation).

Suppose one has a limit function  $f : \Omega \rightarrow {}^*\mathbb{C}$  on a limit set  $\Omega$ . If one asserts that  $f(x) = O(1)$  for each  $x \in \Omega$ , one may be concerned that this statement provides no uniformity in  $x$ . However, it turns out such uniformity is automatic for limit functions.

**Lemma A.5** (Automatic uniformity). *Let  $D \in \mathbb{N}^+$ , and let  $f : \Omega \rightarrow {}^*\mathbb{C}^D$  be a limit function on a limit set  $\Omega$ . Then the following statements are equivalent:*

- (Pointwise boundedness) *For each  $x \in \Omega$ , one has  $f(x) \in \overline{\mathbb{C}}^D$  (i.e.  $f(x) = O(1)$  for all  $x \in \Omega$ ).*
- (Uniform boundedness) *There exists a standard real  $C$  such that  $|f(x)| \leq C$  for all  $x \in \Omega$ .*

Intuitively, this lemma is asserting that the only types of functions that always map unbounded sequences to bounded sequences (but with a bound possibly depending on the initial sequence) are those functions that are uniformly bounded. The lemma can clearly fail if one considers functions  $f$  that are not limit functions; thus it will be important to establish the limit nature of various functions in the arguments below. This lemma is also closely related to the *overspill principle* in nonstandard analysis, or the model-theoretic fact that ultraproducts are countably saturated.

*Proof.* Clearly uniform boundedness implies pointwise boundedness, so we show the converse. Suppose for contradiction that  $f$  was pointwise bounded but not uniformly bounded. Then for every standard integer  $M$  there exists an element  $x_M$  in  $\Omega$  such that  $|f(x_M)| > M$ .

Write  $\Omega$  as the ultralimit of standard sets  $\Omega_{\mathbf{n}}$ , write  $f$  as an ultralimit of a sequence  $f_{\mathbf{n}} : \Omega_{\mathbf{n}} \rightarrow \mathbb{C}^D$ , and write  $x_M = x_{M,\mathbf{n}} \in \Omega_{\mathbf{n}}$ . Thus for each standard  $M$ , the statement  $|f_{\mathbf{n}}(x_{M,\mathbf{n}})| > M$  is true for  $\mathbf{n}$  sufficiently close to  $p$ .

Now we diagonalise. Set  $y = \lim_{\mathbf{n} \rightarrow p} y_{\mathbf{n}}$ , where  $y_{\mathbf{n}} := x_{\mathbf{n},\mathbf{n}}$ . Then  $y \in \Omega$  and one sees that for every standard  $M$ , the statement  $|f_{\mathbf{n}}(y_{\mathbf{n}})| > M$  holds for  $\mathbf{n}$  sufficiently close to  $p$ , thus  $f(y)$  is unbounded. But this contradicts pointwise boundedness.  $\square$

We observe a useful corollary to Lemma A.5.

**Corollary A.6** (Automatic uniform lower bounds). *Let  $D \in \mathbb{N}^+$ , and let  $f : \Omega \rightarrow {}^*\mathbb{C}^D$  be a limit function on a limit set  $\Omega$  such that  $|f(x)| \gg 1$  for all  $x \in \Omega$ . Then there exists a standard  $c > 0$  such that  $|f(x)| \geq c$  for all  $x \in \Omega$ .*

*Proof.* Apply Lemma A.5 to  $1/|f|$ .  $\square$

Inspired by Lemma A.5, we shall simply call a limit function  $f : \Omega \rightarrow {}^*\mathbb{C}^D$  *bounded* if it is either pointwise bounded or uniformly bounded. The space of all bounded limit functions from  $\Omega$  to  ${}^*\mathbb{C}^D$  will be denoted  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D)$ , and we also



write

$$L^\infty(\Omega) = L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega) := \bigcup_{D \in \mathbb{N}^+} L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D). \quad (\text{A.1})$$

When  $D = 1$ ,  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}})$  is a  $*$ -algebra over the bounded complex numbers  $\overline{\mathbb{C}}$  (i.e. it is closed under addition, pointwise multiplication, complex conjugation, and multiplication by bounded complex numbers). It is not, however, a limit set.

For higher dimensions  $D > 1$ , we still have the operations of addition, complex conjugation (conjugating each coefficient of  $\mathbb{C}^D$  separately), and multiplication by bounded complex numbers. However, we do not have a natural product on  $\mathbb{C}^D$ . Instead, we will use the *tensor product*  $\otimes : \mathbb{C}^D \times \mathbb{C}^{D'} \rightarrow \mathbb{C}^{DD'}$ , defined in §3. This induces a tensor product

$$\otimes : L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D) \times L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^{D'}) \rightarrow L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^{DD'})$$

for any  $\Omega$ , which is then a bilinear operation on  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ . Strictly speaking, this tensor product is neither commutative nor associative. However, it is “essentially” commutative and associative in the following sense. Let us say that a function  $f \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D)$  is a *bounded linear combination* of another function  $f' \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^{D'})$  if there exists a linear transformation  $T : {}^*\mathbb{C}^{D'} \rightarrow {}^*\mathbb{C}^D$  with bounded coefficients such that  $f = T \circ f'$ . Then it is clear that for any  $f_1, f_2, f_3 \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ , we have that  $f_2 \otimes f_1$  is a bounded linear combination of  $f_1 \otimes f_2$ , and that  $f_1 \otimes (f_2 \otimes f_3)$  is a bounded linear combination of  $(f_1 \otimes f_2) \otimes f_3$ . This will be a satisfactory substitute for commutativity and associativity for our purposes.

We define the spheres

$$\overline{S^{2D-1}} := \{z \in \overline{\mathbb{C}}^D : |z| = 1\}$$

and

$$\overline{S^\omega} := \bigcup_{D \in \mathbb{N}^+} \overline{S^{2D-1}} = \{z \in \overline{\mathbb{C}}^\omega : |z| = 1\}$$

and observe that  $\overline{S^\omega}$  is closed under complex conjugation and tensor product, and so  $L^\infty(\Omega \rightarrow \overline{S^\omega})$  is also. Also, observe that for any  $f \in L^\infty(\Omega \rightarrow \overline{S^\omega})$ , 1 is a bounded linear combination of  $f \otimes \overline{f}$ .

When  $\Omega$  is a non-empty limit finite set (e.g.  $\Omega = [N]$  or  $\Omega = [N]^k$  for some positive limit integer  $N$  and some standard  $k \geq 1$ ), we have some additional structures.

**Definition A.7** (Bias and correlation). Let  $\Omega$  be a non-empty limit finite set. Given two functions  $f \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ ,  $g \in L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$ , we say that  $f$  and  $g$  *correlate* if one has

$$|\mathbb{E}_{n \in \Omega} f(n) \otimes \overline{g(n)}| \gg 1,$$

and that  $f$  is *biased* if one has

$$|\mathbb{E}_{n \in \Omega} f(n)| \gg 1,$$

i.e. if  $f$  correlates with 1. We say that  $f$  is *unbiased* if it is not biased. We define the  $L^p$  norms

$$\|f\|_{L^p(\Omega)} := (\mathbb{E}_{n \in \Omega} |f(n)|^p)^{1/p}$$

for  $1 \leq p < \infty$ , with the usual convention

$$\|f\|_{L^\infty(\Omega)} := \sup_{n \in \Omega} |f(n)|;$$

these are bounded limit non-negative numbers.

We will also find the following notation useful.

**Definition A.8** (Density). We say that a limit subset  $H$  of a limit finite set  $X$  is *dense* if  $|H| \gg |X|$ , and that a statement  $P(x)$  is true for *many*  $x \in X$  if it is true for all  $x$  in a dense subset  $H$  of  $X$ . If instead  $|H| = o(|X|)$ , we say that  $H$  is a *sparse* subset of  $X$ , and if  $P(x)$  only holds true for  $x$  in a sparse set, we say that  $P(x)$  only holds for *few*  $x \in X$ . If the complement of  $H$  in  $X$  is sparse, we say that  $H$  is a *co-sparse* subset of  $X$ , and if  $P(x)$  holds for all  $x$  in a co-sparse subset, we say that  $P(x)$  holds for *almost all*  $x \in X$ .

A function  $f : X \rightarrow {}^*\mathbb{C}^D$  is said to be *almost bounded* if  $f(x) \in \overline{\mathbb{C}}^D$  for almost all  $x \in X$ . (For instance, for an unbounded limit natural number  $N$ , the function  $n \mapsto \frac{N}{n+1}$  is almost bounded on  $[N]$ .)

*Remarks.* Note that the statement  $P$  does not need to be a limit statement (i.e. the set  $\{x \in X : P(x) \text{ true}\}$  need not be a limit set) for these definitions to make sense; for instance, for  $P$  to hold for many  $x$ , it suffices that  $\{x \in X : P(x) \text{ true}\}$  contain an dense limit subset of  $X$ , but need not be a limit set itself. If one property  $P(x)$  holds for almost all  $x \in X$ , and another property  $Q(x)$  holds for many  $x \in X$ , then  $P(x)$  and  $Q(x)$  simultaneously hold for many  $x \in X$ . However, if  $P$  only holds for many  $x$  rather than for almost all  $x$ , then it need not be the case that  $P(x)$  and  $Q(x)$  simultaneously hold for any  $x$ .

From the pigeonhole principle we see that if a limit set is partitioned into a bounded number of limit pieces, then at least one of the pieces is dense. We can strengthen this principle as follows.

**Lemma A.9** (Pigeonhole principle). *Let  $X$  be a limit finite set, and let  $f$  be an almost bounded limit function from  $X$  to  ${}^*\mathbb{N}$ . Then there exists a dense subset of  $X$  on which  $f$  is constant and equal to a standard natural.*

*Proof.* By hypothesis,  $f$  is bounded on almost all of  $X$ , and hence uniformly bounded on almost all of  $X$  by Lemma A.5. The claim now follows from the pigeonhole principle.  $\square$

We also record here a technical lemma regarding correlation.

**Definition A.10** ( $\sigma$ -limit). A subset  $S$  of an limit set  $X$  is said to be a  $\sigma$ -*limit* set if there is a limit sequence  $n \mapsto S_n$  from limit natural numbers  $n \in {}^*\mathbb{N}$  of limit subsets  $S_n$  of  $X$ , such that  $S$  is the union of the  $S_n$  over all *standard* natural numbers.

*Example.* If  $\Omega$  is a limit set and  $D \in \mathbb{N}^+$ , then the space  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^D)$ , which is an external (i.e. non-limit) subset of the limit space of all limit functions from  $\Omega$  to  ${}^*\mathbb{C}^D$ , is a  $\sigma$ -limit space, since one can express this space as the union, over all standard  $M$ , of the functions bounded uniformly in magnitude by  $M$ . Similarly,  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$  is also a  $\sigma$ -limit set.

**Lemma A.11** (Limit selection lemma). *Let  $X, Y$  be limit sets, let  $R \subset X \times Y$  be a limit relation between  $X$  and  $Y$ , and let  $S$  be a  $\sigma$ -limit subset of  $Y$ . Suppose that for every  $x \in X$  there exists  $s_x \in S$  such that  $(x, s_x) \in R$ . Then there exists a limit function  $x \mapsto s_x$  from  $X$  to  $S$  such that  $(x, s_x) \in R$  for all  $x \in X$ .*

*Remark.* The key point here is the limit nature of the assignment  $x \mapsto s_x$ ; for external (i.e. non-limit) assignments, the claim is immediate from the axiom of choice. There is a similar need for such “measurable selection lemmas” in the ergodic theory analogue of the inverse conjectures for the Gowers norms, see e.g. [36, Appendix A] or [5, Lemma C.4].

*Proof.* We may assume that the sets  $S_n$  in Definition A.10 are increasing in  $n$ .

For each  $x \in X$ , let  $n_x$  be the first limit natural number such that  $(x, s) \in R$  for some  $s \in S_{n_x}$ . By construction,  $x \mapsto n_x$  is a limit map from  $X$  to  ${}^*\mathbb{N}$  which is pointwise bounded. Thus, by Lemma A.5,  $n_x$  is uniformly bounded by some standard natural number  $n_*$ , thus for every  $x \in R$  the set  $\{s \in S_{n_*} : (x, s) \in R\}$  is non-empty. Applying a limit choice function, we may thus find a limit map  $x \mapsto s_x$  with the stated properties.  $\square$

We isolate a special case of this lemma.

**Corollary A.12.** *Let  $\Omega$  be a non-empty limit-finite set, Let  $S$  be a  $\sigma$ -limit subset of  $L^\infty(\Omega \rightarrow \overline{\mathbb{T}}^\omega)$ , and let  $(f_h)_{h \in H}$  be a limit family of limit functions  $f_h \in L^\infty(\Omega \rightarrow \overline{\mathbb{T}}^\omega)$  indexed by an limit set  $H$ , and suppose that for each  $h \in H$ ,  $f_h$  correlates with an element of  $S$ . Then one can find an limit family  $(\phi_h)_{h \in H}$  of functions  $\phi_h \in S$  such that  $f_h$  correlates with  $\phi_h$  for all  $h \in H$ .*

*Proof.* Write  $S$  as the union of limit sets  $S_n$  for standard  $n$ , and let  $S' := \bigcup_{n \in \mathbb{N}} S_n \cup \{n\}$ . Note that this is a  $\sigma$ -limit subset of  $L^\infty(\Omega \rightarrow \overline{\mathbb{T}}^\omega) \times {}^*\mathbb{N}$ . Defining a relation  $R$  between  $H$  and  $S'$  by declaring  $(h, (\phi, n)) \in R$  if  $|\mathbb{E}_{n \in \Omega} f_h(n) \otimes \overline{\phi(n)}| \geq 1/n$ , and applying Lemma A.11, we obtain the claim.  $\square$

## APPENDIX B. POLYNOMIAL ALGEBRA

In section §6 we introduced the notion of a polynomial map between  $I$ -filtered groups  $H$  and  $G$  when the group  $H$  was abelian (Definition 6.18). In this appendix we study the more general notion of a polynomial map, no longer restricting to the case  $H$  abelian. The concept of a polynomial map between groups was introduced by Leibman in [43, 44], and here we adapt it to filtered groups.

Recall the definitions of an ordering  $I$  and of an  $I$ -filtration of a group  $G$  in Definitions 6.7 and 6.8.

**Definition B.1** (Polynomial map). Let  $G, H$  be groups with  $I$ -filtrations  $G_I, H_I$ . If  $g : H \rightarrow G$  is a map then we define the *derivative*  $\partial_h g : H \rightarrow G$  by the formula

$$\partial_h g(n) := g(hn)g(n)^{-1}$$

for all  $n \in H$ . We say that map  $g : H \rightarrow G$  is *polynomial* if one has

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_{i_1 + \dots + i_m}$$

whenever  $m \geq 0$ ,  $i_1, \dots, i_m \in I$ ,  $h_j \in H_{i_j}$  for  $j = 1, \dots, m$  and  $n \in H_0$ . The space of all polynomial maps is denoted  $\text{poly}(H_I \rightarrow G_I)$ .

*Remark.* As mentioned in §3, if  $G$  or  $H$  are written as additive groups instead of multiplicative ones, the definition of partial derivative is adjusted appropriately.

*Example 1.* If  $I = \mathbb{N}$ , and  $H$  is abelian and is given the filtration  $H_i = H$  for  $i = 0, 1$  and  $H_i = \{0\}$  for  $i > 0$ , then a map  $g : H \rightarrow G$  lies in  $\text{poly}(H_{\mathbb{N}}, G_{\mathbb{N}})$  if and only if

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_m$$

for all  $m \geq 0$  and  $h_1, \dots, h_m \in H$ . This coincides with the definition given in [24, Definition 6.1]. Definition B.1 may be considered as a generalisation of this in which the domain group  $H$  is allowed to have nontrivial filtrations.

*Example 2.* Any map  $\phi : G \rightarrow H$  between two  $I$ -filtered groups which is constant and takes values in  $H_0$  is polynomial.

*Example 3.* If  $\phi : H_I \rightarrow G_I$  is a homomorphism of  $I$ -filtered groups that maps  $H_i$  to  $G_i$  for each  $i \in I$ , then  $\phi$  is a polynomial map since, for each  $h \in H$ ,  $\partial_h \phi$  is the constant map  $n \mapsto \phi(h)$ . We will call such homomorphisms  *$I$ -filtered homomorphisms* from the  $I$ -filtered group  $H_I$  to the  $I$ -filtered group  $G_I$ .

*Example 4.* If  $G$  is an  $I$ -filtered group, and  $g \in G$ , then the left translation maps  $x \mapsto gx$  lie in  $\text{poly}(G_I \rightarrow G_I)$ . Indeed, the derivative of this map in any direction  $h \in G_i$  is simply the constant map  $ghg^{-1}$ , which lies in  $G_i$ , and any further derivative of this map is trivial. This example is a special case of the Lazard-Leibman theorem (Corollary B.4 below), since the translation map is the product of a constant map and the identity homomorphism.

*Example 5.* Given three  $I$ -filtered groups  $H, G, G'$ , a map  $g : H \rightarrow G \times G'$  is polynomial ( $G \times G'$  is given the product filtration) if and only if its projections to  $G$  and  $G'$  are polynomial. In other words, we have a canonical isomorphism

$$\text{poly}(H_I \rightarrow (G \times G')_I) \equiv \text{poly}(H_I \rightarrow G_I) \times \text{poly}(H_I \rightarrow G'_I).$$

**HOST-KRA CUBE GROUPS.** There is an important alternative characterisation of polynomial maps in terms of *Host-Kra cube groups*, which we now define. The material in this section is a generalisation of [24], and particularly [24, Proposition 6.5], to the context of polynomial maps  $\text{poly}(H_I \rightarrow G_I)$  (these matters were discussed only in the case  $\text{poly}(H \rightarrow G_I)$ ). The Host-Kra groups are the group-theoretic analogue of the Host-Kra spaces  $X^{[k]}$  of a dynamical system  $X$  introduced in [36].

If  $m$  is a natural number, we let  $2^{[m]}$  be the power set of  $[m] := \{1, \dots, m\}$ .

**Definition B.2.** Let  $G$  be an  $I$ -filtered group, and let  $i_1, \dots, i_m \in I$ . We define the *Host-Kra cube group*  $\text{HK}^{i_1, \dots, i_m}(G_I)$  to be the subgroup of  $G^{2^{[m]}}$  generated by the elements of the form

$$\iota_{\omega_0}(g_{\omega_0}) := (g_{\omega})_{\omega \subset [m]},$$

where  $\omega_0 \subset [m]$ ,  $g_{\omega_0} \in G_{\sum_{j \in \omega_0} i_j}$ , and  $g_{\omega}$  equals  $g_{\omega_0}$  when  $\omega \supseteq \omega_0$  and is the identity otherwise. Thus we see that the  $\iota_{\omega_0}$  are embeddings of  $G_{\sum_{j \in \omega_0} i_j}$  into  $\text{HK}^{i_1, \dots, i_m}(G)$ . We refer to  $m$  as the *order* of the Host-Kra cube groups, and refer to elements of  $\text{HK}^{i_1, \dots, i_m}(G)$  as *cubes of dimension  $m$  and degrees  $i_1, \dots, i_m$* .

*Example.* Let  $G$  be a  $k$ -step nilpotent group, and let  $G_i = [G, G_{i-1}]$  be the lower central series filtration. Then  $\text{HK}^{1, \dots, 1}(G)$  is the subgroup of  $G^{2^{[m]}}$  generated by the “side” elements  $(g_{\omega}^i)_{\omega \subset [m]}$  where  $g_{\omega}^i = g$  if  $i \in \omega$  and  $g_{\omega}^i = \text{id}$  otherwise, for  $i = 1, \dots, m$ , and by the diagonal elements  $(g, \dots, g)$ .

**Theorem B.3.** Let  $G, H$  be  $I$ -filtered groups, and let  $g : H \rightarrow G$  be a map. Then  $g$  is a polynomial map if and only if it preserves cubes, in the sense for every  $m \geq 0$  and  $i_1, \dots, i_m \in I$ , the homomorphism  $g^{2^{[m]}} : H^{2^{[m]}} \rightarrow G^{2^{[m]}}$  defined by

$$g^{2^{[m]}}((h_{\omega})_{\omega \subset [m]}) := (g(h_{\omega}))_{\omega \subset [m]}$$

maps  $\text{HK}^{i_1, \dots, i_m}(H_I)$  to  $\text{HK}^{i_1, \dots, i_m}(G_I)$ .

*Proof.* For inductive reasons it is convenient to establish the following slightly stronger result. For any  $m_0 \geq 0$ , we say that a map  $g : H \rightarrow G$  is *polynomial to order  $m_0$*  if we have

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_{i_1 + \dots + i_m}$$

for all  $m$  with  $0 \leq m \leq m_0$ , all  $i_1, \dots, i_m \in I$ , all  $h_j \in H_{i_j}$  for  $j = 1, \dots, m$ , and all  $n \in H_0$ . It will suffice to show that a map  $g : H \rightarrow G$  is polynomial to order  $m_0$  if and only if it preserves the cubes of dimension up to  $m_0$ .

We establish this by induction on  $m_0$ . The case  $m_0 = 0$  is easy:  $g$  is polynomial to order 0 if it maps  $H_0$  to  $G_0$ , but these are also essentially the Host-Kra groups of order 0, and the claim follows. Now suppose inductively that  $m_0 \geq 1$  and that the claim has already been shown for all smaller values of  $m_0$ .

Suppose first that  $g : H \rightarrow G$  preserves all cubes of dimension up to  $m_0$ . Then by the preceding discussion,  $g$  maps  $H_0$  to  $G_0$ . To show that  $g$  is polynomial to order  $m_0$ , it thus suffices to show that for every  $i \in I$  and  $h \in H_i$ ,  $\partial_h g$  is polynomial to order  $m_0 - 1$  in the shifted  $I$ -filtration  $G_I^{+i}$  defined by

$$G_I^{+i} := (G_{j+i})_{j \in I}. \quad (\text{B.1})$$

By the induction hypothesis, it suffices to show that  $\partial_h g$  preserves cubes of dimension  $m_0 - 1$ . Accordingly, let  $\vec{h} = (h_\omega)_{\omega \subset [m_0-1]}$  be an element of  $\text{HK}^{i_1, \dots, i_{m_0-1}}(H)$ . We may view  $(\vec{h}, h \cdot \vec{h})$  as an element of  $\text{HK}^{i_1, \dots, i_{m_0-1}, i}(H)$  of one higher order, where  $h \cdot \vec{h} := (hh_\omega)_{\omega \subset [m_0-1]}$ . By hypothesis on  $g$ , we have

$$(g^{2^{[m_0-1]}}(\vec{h}), g^{2^{[m_0-1]}}(h \cdot \vec{h})) \in \text{HK}^{i_1, \dots, i_{m_0-1}, i}(G).$$

An inspection of Definition B.2 reveals that  $(\vec{g}_1, \vec{g}_2)$  lies in  $\text{HK}^{i_1, \dots, i_{m_0-1}, i}(G)$  if and only if  $\vec{g}_1$  lies in  $\text{HK}^{i_1, \dots, i_{m_0-1}}(G)$  and  $\vec{g}_2(\vec{g}_1)^{-1}$  lies in  $\text{HK}^{i_1, \dots, i_{m_0-1}}(G, G_I^{+i})$  (which is easily seen to be a normal subgroup of  $\text{HK}^{i_1, \dots, i_{m_0-1}}(G)$ ). We conclude that

$$g^{2^{[m_0-1]}}(h \cdot \vec{h}) \cdot g^{2^{[m_0-1]}}(\vec{h})^{-1} \in \text{HK}^{i_1, \dots, i_{m_0-1}}(G, G_I^{+i}).$$

But

$$g^{2^{[m_0-1]}}(h \cdot \vec{h}) \cdot g^{2^{[m_0-1]}}(\vec{h})^{-1} = (\partial_h g)^{2^{[m_0-1]}}(\vec{h}),$$

and the claim follows.

Next, suppose conversely that  $g : H \rightarrow G$  is a polynomial map of order up to  $m_0$ ; by the inductive hypothesis, it suffices to show that  $g$  preserves all the cubes of dimension exactly  $m_0$ . Accordingly, let  $\vec{h}$  be an element of  $\text{HK}^{i_1, \dots, i_{m_0}}(H)$  of this dimension. Arguing as before, we may write

$$\vec{h} = (\vec{h}_1, \vec{h}_2 \vec{h}_1)$$

where  $\vec{h}_1 \in \text{HK}^{i_1, \dots, i_{m_0-1}}(H)$  and  $\vec{h}_2 \in \text{HK}^{i_1, \dots, i_{m_0-1}}(H, H_I^{+i_{m_0}})$ . Our objective is then to show that

$$g^{2^{[m_0]}}(\vec{h}) = (g^{2^{[m_0-1]}}(\vec{h}_1), g^{2^{[m_0-1]}}(\vec{h}_2 \vec{h}_1))$$

lies in  $\text{HK}^{i_1, \dots, i_{m_0}}(G)$ . By the decomposition of  $\text{HK}^{i_1, \dots, i_{m_0}}(G)$ , it thus suffices to show that

$$g^{2^{[m_0-1]}}(\vec{h}_2 \vec{h}_1) g^{2^{[m_0-1]}}(\vec{h}_1)^{-1} \in \text{HK}^{i_1, \dots, i_{m_0-1}}(G, G_I^{+i_{m_0}}). \quad (\text{B.2})$$

Recall that  $\text{HK}^{i_1, \dots, i_{m_0-1}}(H, H_I^{+i_{m_0}})$  is generated by elements of the form  $\iota_{\omega_0}(h_{\omega_0})$ , where  $\omega_0 \subset [m_0 - 1]$  and  $h_{\omega_0} \in H_{\sum_{j \in \omega_0} i_j + i_{m_0}}$ . By telescoping series, we thus see

that to establish the above claim it suffices to do so under the additional assumption that  $\vec{h}_2$  is a generator

$$\vec{h}_2 = \iota_{\omega_0}(h_{\omega_0})$$

for some  $\omega_0 \subset [m-1]$  and  $h_{\omega_0} \in H_{\sum_{j \in \omega_0} i_j + i_{m_0}}$ .

By relabeling we may assume that  $\omega_0 = \{m'+1, \dots, m_0-1\}$  for some  $0 \leq m' \leq m_0-1$ . The left-hand side of (B.2) then simplifies to

$$(\partial_{h_{\omega_0}} g)^{2^{[m']}}(\vec{h}'_1), \quad (\text{B.3})$$

where  $\vec{h}'_1$  is the restriction of  $\vec{h}_1$  to  $2^{[m']}$ , and we embed  $G^{2^{[m']}}$  into  $G^{2^{[m_0-1]}}$  by identifying  $(g_\omega)_{\omega \subset [m']}$  with the tuple  $(\tilde{g}_\omega)_{\omega \subset [m_0-1]}$ , where  $\tilde{g}_\omega$  is equal to  $g_{\omega \cap [m']}$  when  $\omega$  contains  $\Omega$ , and is equal to the identity otherwise.

But by induction hypothesis, (B.3) lies in  $\text{HK}^{i_1, \dots, i_{m_0-1}}(G, G_I^{+\sum_{j \in \omega_0} i_j + i_{m_0}})$ . By Definition B.2, this embeds into  $\text{HK}^{i_1, \dots, i_{m_0-1}}(G, G_I^{+i_{m_0}})$ , giving (B.2) as desired, and the claim follows.  $\square$

Theorem B.3 has two immediate corollaries.

**Corollary B.4** (Lazard-Leibman theorem). *Let  $G, H$  be  $I$ -filtered groups. Then  $\text{poly}(H_I \rightarrow G_I)$  is also a group (using pointwise multiplication as a group operation).*

**Corollary B.5** (Composition). *Let  $G, H, K$  be  $I$ -filtered groups. If  $g \in \text{poly}(H_I \rightarrow G_I)$  and  $h \in \text{poly}(K_I \rightarrow H_I)$ , then  $g \circ h \in \text{poly}(H_I \rightarrow K_I)$ .*

In other words, for any fixed  $I$ , the class of  $I$ -filtered groups together with their polynomial maps form a category. It is remarkably difficult to establish Corollary B.5 in full generality without the machinery of Host-Kra cube groups.

*Example.* If  $G, H$  are  $I$ -filtered groups with  $H = (H, +)$  abelian, and  $g$  is a polynomial map from  $H$  to  $G$ , then the translates  $g(\cdot + h)$  and dilates  $g(q \cdot)$  for  $h \in H$  and  $q \in \mathbb{Z}$  are also polynomial maps from  $H$  to  $G$ , thanks to Corollary B.5 and Examples 3 and 4 following Definition B.1. More generally, if  $\phi : H' \rightarrow H$  is a filtered homomorphism and  $g \in \text{poly}(H_I \rightarrow G_I)$ , then  $g \circ \phi \in \text{poly}(H'_I \rightarrow G_I)$ .

*Example.* Using Corollary B.4 we can establish that any algebraic word  $w$  on  $k$  generators defines a polynomial map from  $H^k$  to  $H$  for any  $I$ -filtered group  $H$ . For instance, the map  $(g, h) \rightarrow g^2 h^{-3} g h$  is a polynomial map from  $H \times H$  to  $H$ .

We can strengthen Corollary B.4 slightly, by giving  $\text{poly}(H_I \rightarrow G_I)$  the structure of an  $I$ -filtered group:

**Proposition B.6** (Filtered Lazard-Leibman theorem). *Let  $(G, G_I), (H, H_I)$  be  $I$ -filtered groups. Then  $\text{poly}(H_I \rightarrow G_I)$  is also an  $I$ -filtered group, with filtration  $(\text{poly}(H_I \rightarrow G_I^{+i}))_{i \in I}$ , where the shifted filtration  $G_I^{+i}$  was defined in (B.1). In particular, the  $\text{poly}(H_I \rightarrow G_I^{+i})$  are normal subgroups of  $\text{poly}(H_I \rightarrow G_I)$ .*

*Proof.* The only non-trivial claim to show is that if  $g_i \in \text{poly}(H_I \rightarrow G_I^{+i})$  and  $g_j \in \text{poly}(H_I \rightarrow G_I^{+j})$  for some  $i, j \in I$ , then  $[g_i, g_j] \in \text{poly}(H_I \rightarrow G_I^{+i+j})$ . It suffices to show for each  $m_0 \geq 0$  that if  $g_i, g_j$  are polynomial maps up to order  $m_0$  from  $(H, H_I)$  to  $(G, G_I^{+i}), (G, G_I^{+j})$  respectively, then  $[g_i, g_j]$  is a polynomial map up to order  $m_0$  from  $(H, H_I)$  to  $(G, G_I^{+i+j})$ .

Again we induct on  $m_0$ . The case  $m_0 = 0$  is trivial, so suppose that  $m_0 \geq 1$  and that the claim has already been proven for smaller values of  $m_0$ .

As  $g_i, g_j$  map  $H_0$  to  $G_i, G_j$  respectively,  $[g_i, g_j]$  maps  $H_0$  to  $G_{i+j}$ . It thus suffices to show that for each  $k \in I$  and  $h \in H_k$ , that  $\partial_h[g_i, g_j]$  is a polynomial map up to order  $m_0 - 1$  from  $(H, H_I)$  to  $(G, G_I^{+i+j+k})$ . But a brief calculation shows that

$$\partial_h[g_i, g_j] = g_i^{-1}(\partial_h g_i)^{-1} g_j^{-1}(\partial_h g_j)^{-1}(\partial_h g_i) g_i (\partial_h g_j) g_i^{-1} g_j g_i. \quad (\text{B.4})$$

By induction hypothesis (and Corollary B.4), the maps that are polynomial up to order  $m_0 - 1$  from  $(H, H_I)$  to  $(G, G_I^{+i+j+k})$  form a normal subgroup of the maps that are polynomial up to order  $m_0 - 1$  from  $(H, H_I)$  to  $(G, G_I)$ . If we quotient out by this normal subgroup, then a further application of the induction hypothesis shows that  $\partial_h g_i$  commutes with  $g_j$  and  $\partial_h g_j$ , and that  $g_i$  commutes with  $\partial_h g_j$ . An inspection of (B.4) then shows that the right-hand side vanishes once one quotients out by this normal subgroup, and the claim follows.  $\square$

Proposition B.6 has some useful corollaries:

**Corollary B.7** (Approximate linearity and commutativity). *Let  $G, H$  be  $I$ -filtered groups, let  $i, j, k, l \in I$ , and let  $g_i \in \text{poly}(H_I \rightarrow G_I^{+i})$ ,  $g_j \in \text{poly}(H_I \rightarrow G_I^{+j})$ ,  $h_k \in H_k$ , and  $h_l \in H_l$ . Then we have*

$$\partial_{h_k}(g_i g_j) = (\partial_{h_k} g_i)(\partial_{h_k} g_j) \bmod \text{poly}(H_I \rightarrow G_I^{+i+j+k}) \quad (\text{B.5})$$

and

$$\partial_{h_k h_l}(g_i) = (\partial_{h_k} g_i)(\partial_{h_l} g_i) \bmod \text{poly}(H_I \rightarrow G_I^{+i+k+l}). \quad (\text{B.6})$$

If  $H$  is abelian, we also have

$$(\partial_{h_l} g_i)(\partial_{h_k} g_i) = (\partial_{h_k} g_i)(\partial_{h_l} g_i) \bmod \text{poly}(H_I \rightarrow G_I^{+i+k+l}). \quad (\text{B.7})$$

*Proof.* The conclusions (B.5), (B.6) follow from Proposition B.6 and the identities

$$\partial_{h_k}(g_i g_j) = (\partial_{h_k} g_i)(\partial_{h_k} g_j)[\partial_{h_k} g_j, g_i^{-1}]$$

and

$$\partial_{h_k h_l}(g_i) = (\partial_{h_l} \partial_{h_k} g_i)(\partial_{h_k} g_i)(\partial_{h_l} g_i). \quad (\text{B.8})$$

The identity (B.7) then follows by swapping the roles of  $h_k$  and  $h_l$  in (B.6).  $\square$

Next, we make the useful observation that in order to check polynomiality of a map, it suffices to do so on generators.

**Proposition B.8** (Checking polynomiality on generators). *Let  $G, H$  be  $I$ -filtered groups. For each  $i \in I$ , let  $E_i$  be a set of generators for  $H_i$ . Then a map  $g : H \rightarrow G$  is polynomial if and only if one has*

$$\partial_{h_1} \dots \partial_{h_m} g(n) \in G_{i_1 + \dots + i_m} \quad (\text{B.9})$$

for all  $m \geq 0$ , all  $i_1, \dots, i_m \in I$ , and all  $h_j \in E_{i_j}$  for  $j = 1, \dots, m$ , and all  $n \in H_0$ .

*Proof.* The “only if” part is trivial, so it suffices to prove the “if” part. For inductive reasons, we shall prove the following more general statement: if  $l, m_0 \geq 0$ , and  $g : H \rightarrow G$  is such that  $\partial_{h_1} \dots \partial_{h_m} g$  is a polynomial map up to order  $l$  from  $(H, H_I)$  to  $(G, G_I^{+i_1 + \dots + i_m})$  whenever  $0 \leq m \leq m_0$ ,  $i_1, \dots, i_m \in I$  and  $h_j \in E_{i_j}$  for  $j = 1, \dots, m$ , then  $g$  is a polynomial map from  $H$  to  $G$  up to order  $m_0 + l$ . Indeed, by setting  $l = 0$  and sending  $m_0 \rightarrow \infty$  we obtain the claim.

We establish the claim by induction on  $m$ . The case  $m_0 = 0$  is trivial, so suppose that  $m_0 \geq 1$  and that the claim has already been proven for smaller values of  $m_0$ .

Fix  $l$ . Let  $1 \leq m \leq m_0$  and  $i_1, \dots, i_m \in I$ , and suppose that  $h_j \in E_{i_j}$  for  $j = 2, \dots, m$ , and write  $\tilde{g} := \partial_{h_2} \dots \partial_{h_m} g$ . By hypothesis, we have that  $\partial_{h_1} \tilde{g}$  is

a polynomial map of order  $l$  from  $(H, H_I)$  to  $(G, G_I^{+i_1+\dots+i_m})$  whenever  $h_1$  lies in  $E_{i_1}$ . Using (B.8) and Corollary B.4, we conclude the same statement holds when  $h_1$  lies in  $H_{i_1}$ . Also, by induction hypothesis  $\tilde{g}$  is also known to be a polynomial map of order  $l$  from  $(H, H_I)$  to  $(G, G_I^{+i_2+\dots+i_m})$ . We conclude that  $\tilde{g}$  is in fact a polynomial map of order  $l+1$  from  $(H, H_I)$  to  $(G, G_I^{+i_2+\dots+i_m})$ . Applying the induction hypothesis again, we conclude that  $g$  is a polynomial map of order  $l+m$  from  $H$  to  $G$ , as required.  $\square$

*Example.* Let  $G_1, G_2, G$  be  $I$ -filtered groups, and let  $B : G_1 \times G_2 \rightarrow G$  be a map which is “bilinear” in the sense that the maps  $g_1 \mapsto B(g_1, g_2)$  for fixed  $g_2 \in G_2$  and  $g_2 \mapsto B(g_1, g_2)$  for fixed  $g_1 \in G_1$  are homomorphisms, and such that  $B$  maps  $G_{1, \geq i} \times G_{2, \geq j}$  to  $G_{i+j}$  for any  $i, j \in I$ . Then  $B$  is a polynomial map, as can be seen by using Proposition B.8 with  $G_{1, \geq i} \times \{\text{id}\} \cup \{\text{id}\} \times G_{2, \geq i}$  as the generating set for  $(G_1 \times G_2)_i = G_{1, \geq i} \times G_{2, \geq i}$ . Combining this with Corollary B.5, we conclude in particular that if  $H$  is an  $I$ -filtered group and  $g_1 \in \text{poly}(H_I \rightarrow (G_1)_I)$ ,  $g_2 \in \text{poly}(H_I \rightarrow (G_2)_I)$ , then  $B(g_1, g_2) \in \text{poly}(H_I \rightarrow G_I)$ ; informally, this is asserting that the product of polynomials is again a polynomial.

*Example.* Let  $G$  be an  $\mathbb{N}^k$ -filtered group, and let  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow G_{\mathbb{N}^k})$  be a polynomial sequence, in which  $\mathbb{Z}^k$  is given the multidegree filtration. We can collapse the  $\mathbb{N}^k$ -filtration on  $G$  to an  $\mathbb{N}$ -filtration by defining  $G_i$  to be the group generated by  $G_{(i_1, \dots, i_k)}$  for all  $(i_1, \dots, i_k) \in \mathbb{N}^k$  with  $i_1 + \dots + i_k = i$ . From Proposition B.8 we thus conclude that  $g$  remains a polynomial map from  $\mathbb{Z}^k$  to  $G$  if we now give  $\mathbb{Z}^k$  the degree filtration, and give  $G$  the  $\mathbb{N}$ -filtration indicated above.

The next lemma describes a useful type of Taylor expansion for polynomial sequences.

**Lemma B.9** (Taylor expansion). *Let  $d \geq 1$  be a natural number, let  $G$  be an  $\mathbb{N}^d$ -filtered group of degree  $\subset J$  for some finite downset  $J$ , and let  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}^d}^d \rightarrow G_{\mathbb{N}^d})$ , where  $\mathbb{Z}^d$  is given the multidegree filtration. We complete the partial ordering on  $J$  to a total ordering in some arbitrary fashion. Then there exist unique Taylor coefficients  $g_j \in G_j$  for each  $j \in J$  such that*

$$g(n) = \prod_{j \in J} g_j^{\binom{n}{j}}.$$

Here we adopt the notational convention

$$\binom{(n_1, \dots, n_d)}{(j_1, \dots, j_d)} := \binom{n_1}{j_1} \dots \binom{n_d}{j_d}.$$

*Proof.* We first show uniqueness. Suppose that we have two Taylor expansions that agree everywhere, that is to say

$$\prod_{j \in J} g_j^{\binom{n}{j}} = \prod_{j \in J} (g'_j)^{\binom{n}{j}}$$

for all  $n \in \mathbb{Z}^d$ . Setting  $n = 0$  we see that  $g_0 = g'_0$ . Cancelling this, we see that

$$\prod_{j \in J: j > 0} g_j^{\binom{n}{j}} = \prod_{j \in J: j > 0} (g'_j)^{\binom{n}{j}}.$$



More generally, suppose inductively that we have shown that  $g_j = g'_j$  for all  $j \leq j_0$  and

$$\prod_{j \in J: j > j_0} g_j^{(n)} = \prod_{j \in J: j > j_0} (g'_j)^{(n)}$$

for all  $n \in \mathbb{Z}^d$  some  $j_0 \in J$ . If  $j_0$  is the maximal element of  $J$  then we are done. Otherwise, let  $j_1$  be the next element after  $j_0$  in the total ordering of  $J$ . Setting  $n = j_1$  we conclude that  $g_{j_1} = g'_{j_1}$ , and then we can continue the induction and establish uniqueness.

Now we show existence by inducting on the cardinality of  $J$ . The claim is trivial for  $J$  empty, so suppose that  $J$  is non-empty, and let  $j_*$  be the maximal element of  $J$ . The group  $G_{j_*}$  is a central subgroup of  $G$ ; if we quotient  $G$  by  $G_{j_*}$ , we obtain an  $\mathbb{N}^d$ -filtered group  $G/G_{j_*}$  of degree  $\subset J \setminus \{j_*\}$ . Let  $\pi : G \rightarrow G_{j_*}$  be the quotient map. Applying the induction hypothesis, we have a Taylor expansion

$$\pi(g(n)) = \prod_{j \in J: j \neq j_*} h_j^{(n)}$$

for some  $h_j \in \pi(G_j)$ . Writing  $h_j = \pi(g_j)$  for some  $g_j \in G_j$ , and using the central nature of  $G_{j_*}$ , we conclude that

$$g(n) = \left( \prod_{j \in J: j \neq j_*} g_j^{(n)} \right) g'(n)$$

for some  $g'(n)$  taking values in  $G_{j_*}$ . By Corollary B.4,  $g'$  is a polynomial sequence, and therefore

$$\partial_{e_1}^{j_1} \dots \partial_{e_k}^{j_k} g'(n) = \text{id}$$

whenever  $(j_1, \dots, j_k) \not\leq j_*$ , with  $e_1, \dots, e_k$  being the basis of  $\mathbb{Z}^k$ . We can “integrate” this difference equation repeatedly using the abelian nature of  $G_{j_*}$  (and the Pascal’s triangle relation  $\partial_{e_i} \binom{n}{j+e_i} = \binom{n}{j}$ ) and conclude that

$$g'(n) = \prod_{j \leq j_*} (g'_j)^{(n)}$$

for some  $g'_j \in G_{j_*}$ . Using the central nature of  $G_{j_*}$ , we conclude that

$$g(n) = \prod_{j \in J} (g_j g'_j)^{(n)}$$

(with the convention that  $g_{j_*} = \text{id}$ ) and the claim follows.  $\square$

**Corollary B.10** (Pullback). *Let  $d \geq 1$  be a natural number, let  $G$  be an  $\mathbb{N}^d$ -filtered group of degree  $\subset J$  for some finite  $J$ , and let  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}^d}^d \rightarrow G_{\mathbb{N}^d})$ . Let  $\phi : G' \rightarrow G$  be a  $\mathbb{N}^d$ -filtered homomorphism of  $\mathbb{N}^d$ -filtered groups such that  $\phi : G'_j \rightarrow G_j$  is surjective for every  $j$ . Then there exists  $g' \in \text{poly}(\mathbb{Z}_{\mathbb{N}^d}^d \rightarrow G'_{\mathbb{N}^d})$  such that  $g = g' \circ \phi$ .*

*Proof.* Apply Lemma B.9 and then pull back each of the resulting Taylor coefficients  $g_j$ .  $\square$

## APPENDIX C. LIFTING LINEAR NILSEQUENCES TO POLYNOMIAL ONES

The purpose of this appendix is to demonstrate the equivalence of the linear inverse conjecture, Conjecture 1.2, with the polynomial inverse conjecture, Conjecture 4.5. We remind the reader that this is not strictly speaking necessary to establish the results in [23], but the latter paper was written before the more general notion of a polynomial nilsequence came to the fore.

The key observation here is that every polynomial nilsequence of degree  $\leq s$  can be “lifted” to an  $s$ -step linear nilsequence in a certain sense.

We begin by recording a useful lemma.

**Lemma C.1** (Discrete polynomials are cocompact). *Let  $G/\Gamma$  be an  $\mathbb{N}$ -filtered nilmanifold. Then  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  is a lattice (i.e. a discrete cocompact subgroup) of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  (where we give  $\mathbb{Z}$  the degree filtration).*

*Proof.* We may assume that  $G/\Gamma$  has degree-rank  $\leq d$ . It will suffice to show that any polynomial sequence  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  can be factorised as  $g = \gamma g'$  where  $\gamma \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  and  $g'$  ranges in a compact subset of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ . It is enough to show by induction on  $i$  that for every  $0 \leq i \leq d+1$ , there exists a factorisation  $g = \gamma_i h_i g'_i$  where  $\gamma_i \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$ ,  $g'_i$  lies in a compact subset of  $\text{poly}(\mathbb{Z} \rightarrow G)$ , and  $h_i \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  is such that  $h(0) = \dots = h(i-1) = \text{id}$ , since for  $i = d+1$  this forces  $h$  to be trivial.

This inductive claim is trivial for  $i = 0$  (setting  $\gamma_0 = g'_0$  to be trivial). Now suppose inductively that one has a factorisation  $g = \gamma_i h_i g'_i$  for some  $0 \leq i \leq d$ . Since  $h(0) = \dots = h(i-1) = \text{id}$ , we see from Taylor expansion that  $h(i) \in G_i$ . Since  $\Gamma_i := \Gamma \cap G_i$  is cocompact in  $G_i$ , we may factorise  $h(i) = \tilde{\gamma}_{i+1}(i) \tilde{g}'_{i+1}(i)$  for some  $\tilde{\gamma}_{i+1}(i) \in \Gamma_i$  and  $\tilde{g}'_{i+1}(i)$  in a cocompact subset of  $G_i$ . By Taylor expansion we may extend  $\tilde{\gamma}_{i+1}, \tilde{g}'_{i+1}$  to elements of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  and of a compact subset of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  respectively which are trivial on  $0, \dots, i-1$ . Writing  $\gamma_{i+1} := \gamma_i \tilde{\gamma}_{i+1}$ ,  $h_{i+1} := \tilde{\gamma}_{i+1}^{-1} h_i (\tilde{g}'_{i+1})^{-1}$ , and  $g'_{i+1} := \tilde{g}'_{i+1} g'_i$  we obtain the claim.  $\square$

Now we establish the key lifting proposition.

**Proposition C.2** (Polynomial nilsequences can be lifted to linear ones). *Let  $G/\Gamma$  be a filtered nilmanifold of degree  $\leq s$ . Then there exists a standard  $s$ -step nilmanifold  $\tilde{G}/\tilde{\Gamma}$ , a standard compact subset  $K$  of  $\tilde{G}/\tilde{\Gamma}$ , and a standard Lipschitz map  $\pi : K \rightarrow G/\Gamma$ , such that for every (standard) polynomial sequence  $g : \mathbb{Z} \rightarrow G$ , there exists  $\tilde{g} \in \tilde{G}$  and  $\tilde{x} \in \tilde{G}/\tilde{\Gamma}$  such that  $\tilde{g}^n \tilde{x} \in K$  and  $g(n)^* \Gamma = \pi(\tilde{g}^n \tilde{x})$  for all  $n \in \mathbb{Z}$ .*

Indeed, with this proposition, any degree  $\leq s$  nilsequence  $n \mapsto F(g(n)^* \Gamma)$  can then be lifted to an  $s$ -step linear nilsequence  $n \mapsto (F \circ \pi)(\tilde{g}^n \tilde{x})$  with  $\tilde{g} \in {}^* \tilde{G}$  and  $\tilde{x} \in {}^*(\tilde{G}/\tilde{\Gamma})$ , where  $F \circ \pi$  is extended from a Lipschitz function on  $K$  to a Lipschitz function on  $\tilde{G}/\tilde{\Gamma}$  in some arbitrary fashion. From this one easily concludes that Conjecture 1.2 follows from Conjecture 4.5. (The converse implication is trivial, because every linear nilsequence is a polynomial nilsequence.)

To motivate Proposition C.2 let us present an illustrative example. We take  $s = 2$  and  $G/\Gamma$  to be the unit circle  $\mathbb{R}/\mathbb{Z}$  with the quadratic filtration (thus  $G_i$  equals  $\mathbb{R}$  for  $i \leq 2$  and  $\{0\}$  for  $i > 2$ ). By Remark 9.6, a polynomial sequence  $g : \mathbb{Z} \rightarrow G$  then takes the form  $g(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \alpha_2 \binom{n}{2}$  for some frequencies  $\alpha_0, \alpha_1, \alpha_2$  (i.e. a non-standard classical quadratic polynomial). To lift this quadratic sequence to a linear one, we introduce use the Heisenberg nilmanifold  $\tilde{G}/\tilde{\Gamma}$  (Example 6.1), and

place inside it the *skew torus*

$$K := \{g_1^{t_1}[g_1, g_2]^{t_{[1,2]}}\Gamma : t_1, t_{[1,2]} \in \mathbb{R}\}.$$

This is easily seen to be compact (indeed, it is topologically equivalent to  $\mathbb{T}^2$ ). We define the map  $\pi : K \rightarrow \mathbb{T}$  by the formula

$$\pi(g_1^{t_1}[g_1, g_2]^{t_{[1,2]}}) := t_{[1,2]} \bmod 1;$$

it is easy to see that  $\pi$  is well-defined and smooth. If we set

$$\tilde{g} := g_1^\alpha g_2[g_1, g_2]^\beta; \quad \tilde{x} := [g_1, g_2]^\gamma \tilde{\Gamma}$$

for some frequencies  $\alpha, \beta, \gamma \in \mathbb{R}$ , then a brief calculation shows that for any integer  $n$ ,  $\tilde{g}^n \tilde{x}$  lies in  $K$  and

$$\pi(\tilde{g}^n \tilde{x}) = \frac{n(n+1)}{2}\alpha + n\beta + \gamma \bmod 1,$$

and so one can arrange for  $\pi(\tilde{g}^n \tilde{x}) = g(n)$  by choosing  $\alpha, \beta, \gamma$  appropriately in terms of  $\alpha_0, \alpha_1, \alpha_2$ .

The above construction was *ad hoc* in nature, requiring one to conjure up the Heisenberg group out of thin air. However, it is possible to canonically construct a lifted nilmanifold  $\tilde{G}/\tilde{\Gamma}$  in the general case. Fix  $G/\Gamma$ . By Remark 9.6,  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  is a Lie group topologically isomorphic to  $\prod_{i \geq 0} G_i$ , but with a different group structure. Since  $G$  has degree  $< s+1$ , we see that  $G$  is  $\leq s$ -step nilpotent, which implies that  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  is  $\leq s$ -step nilpotent also.

Let  $\Gamma_{\mathbb{N}}$  be the restriction of the filtration  $G_{\mathbb{N}}$  to  $\Gamma$  (Example 6.14), thus  $\Gamma$  is now a filtered group. By Lemma C.1,  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  has the structure of an  $s$ -step nilmanifold. This is not yet the nilmanifold  $\tilde{G}/\tilde{\Gamma}$  needed for Proposition C.2, but we can modify it as follows. We observe that there is a shift automorphism  $T$  acting on both  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  and  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  by the formula  $Tg(n) := g(n+1)$ . It also acts on the Lie algebra  $\log \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ , which by abuse of notation we shall call  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$ . This action is unipotent; indeed,  $T-1$  maps  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}}^{+i})$  to  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}}^{+(i+1)})$  for all  $i \geq 0$ , where  $G^{+i}$  is  $G$  with the shifted filtration  $G_d^{+i} := G_{d+i}$ . The conjugation action of  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$  has the same unipotence property by the filtered nature of  $G$ . Because of this, we see that the conjugation action of semi-direct product<sup>9</sup>  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}}) \rtimes_T \mathbb{Z}$  on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$  is  $s$ -step unipotent, which implies that  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}}) \rtimes_T \mathbb{Z}$  is  $s$ -step nilpotent.

Unfortunately, the group  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}}) \rtimes_T \mathbb{Z}$  is not connected, so it is not directly suitable for the purposes of establishing Proposition C.2. But this can be easily remedied by using the unipotent nature of the action of  $T$  on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$  to express<sup>10</sup>  $T = T^1$  for some smooth unipotent group action  $t \mapsto T^t$  of the real line  $\mathbb{R}$  on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$ , which can then be exponentiated to provide a unipotent group action (which we will also call  $t \mapsto T^t$ ) on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ . The action of the group  $\tilde{G} := \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}}) \rtimes_T \mathbb{R}$  on  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \log G_{\mathbb{N}})$  is then  $s$ -step unipotent, which implies that  $\tilde{G}$  is  $s$ -step nilpotent.

<sup>9</sup>Note that  $\mathbb{Z}$  is viewed as an additive group, while  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  is viewed as a multiplicative group; we hope that this will not cause confusion.

<sup>10</sup>This can also be done by the machinery of Mal'cev bases for both discrete and continuous nilpotent groups, see [45].

The group  $\tilde{G}$  is an  $s$ -step nilpotent Lie group which is both connected and simply connected. It contains the discrete subgroup  $\tilde{\Gamma} := \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}}) \rtimes_T \mathbb{Z}$ . Since  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$  is cocompact in  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$  (and  $\mathbb{Z}$  is cocompact in  $\mathbb{R}$ ),  $\tilde{G}$  is cocompact in  $\tilde{G}$ ; thus  $\tilde{G}/\tilde{\Gamma}$  has the structure of a nilmanifold.

There is a canonical map  $\theta$  from  $\tilde{G}/\tilde{\Gamma}$  to  $\mathbb{T}$  induced by the projections of  $\tilde{G}, \tilde{\Gamma}$  to  $\mathbb{R}$  and  $\mathbb{Z}$  respectively. We denote the kernel  $\theta^{-1}(\{0\})$  of this map by  $K$ , thus  $K$  is a compact subset of  $\tilde{G}, \tilde{\Gamma}$ . Observe that every element of  $K$  can be represented as  $(g, 0)\tilde{\Gamma}$  for some  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ , which is unique up to multiplication on the right by  $\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow \Gamma_{\mathbb{N}})$ . We then define the map  $\pi : K \rightarrow G/\Gamma$  by the formula  $\pi(g) := g(0)\Gamma$ ; it is clear that  $\pi$  is a Lipschitz continuous map.

We are now ready to establish Proposition C.2. Let  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ , then we set  $\tilde{x} := (g, 0)\tilde{\Gamma} \in K$  and  $\tilde{g} := (\text{id}, 1) \in \tilde{G}$ . One easily verifies that for any integer  $n$ ,  $\tilde{g}^n \tilde{x} = (T^n g, 0)\tilde{\Gamma} \in K$ , and so  $\pi(\tilde{g}^n \tilde{x}) = g(n)$ . Proposition C.2 follows.

#### APPENDIX D. EQUIDISTRIBUTION THEORY

The purpose of this appendix develop the quantitative Ratner-type equidistribution theory for nilmanifolds, which will help us determine when averages such as

$$\mathbb{E}_{n \in [N]} F(\mathcal{O}(n)) \quad (\text{D.1})$$

are large, for various nilsequences  $n \mapsto F(\mathcal{O}(n))$ . We will also need a multidimensional version<sup>11</sup> of this theory, in which  $[N]$  is replaced with  $[N]^k$ , or more generally by the Cartesian product of  $k$  arithmetic progressions.

This theory is based on the results [24] on equidistribution in nilmanifolds, translated to the language of ultralimits. The results in this appendix will be needed in two places. Firstly, Theorem D.6 below, which gives a criterion for when averages such as (D.1) are large, will be used in §11 to analyse the correlation property arising from Proposition 7.3. Secondly, Theorem D.5, which (locally) factorises an arbitrary multidimensional polynomial orbit into equidistributed and smooth pieces, will be used to give an important criterion for when a nilcharacter is biased (see Lemma E.11).

We begin with some basic definitions.

**Definition D.1** (Equidistribution). Let  $G/\Gamma$  be a standard nilmanifold, which then admits a canonical Haar probability measure  $\mu$ . Let  $\Omega$  be a non-empty limit finite set, and let  $\mathcal{O} : \Omega \rightarrow {}^*(G/\Gamma)$  be a limit function. We say that  $\mathcal{O}$  is *equidistributed* in  $G/\Gamma$  if, for every  $F \in \text{Lip}(G/\Gamma)$ , one has

$$\mathbb{E}_{n \in \Omega} F(\mathcal{O}(n)) = \int_{G/\Gamma} F d\mu + o(1), \quad (\text{D.2})$$

or equivalently if  $n \mapsto F(\mathcal{O}(n))$  is unbiased on  $\Omega$  whenever  $\int_{G/\Gamma} F d\mu = 0$ .

Now we specialise to the case  $\Omega = [N]^k$ . We say that  $\mathcal{O}$  is *totally equidistributed* on  $[N]^k$  if it is equidistributed on every product  $P_1 \times \dots \times P_k$  of dense arithmetic progressions  $P_1, \dots, P_k$  in  $[N]$ , thus

$$\mathbb{E}_{n \in P_1 \times \dots \times P_k} F(\mathcal{O}(n)) = \int_{G/\Gamma} F d\mu + o(1) \quad (\text{D.3})$$

for every standard Lipschitz function  $F : G/\Gamma \rightarrow \mathbb{C}$ .

<sup>11</sup>On the other hand, we will however only need to work with the degree filtration, although it is certain that the theory here would extend to  $I$ -filtered nilsequences for other orderings.

*Remark.* We defined equidistribution using standard Lipschitz functions  $F \in \text{Lip}(G/\Gamma)$ , but the statement (D.2) for  $F \in \text{Lip}(G/\Gamma)$  automatically implies the same claim for  $F \in \text{Lip}^*(G/\Gamma)$ .

This notion of equidistribution on  $[N]$  is closely related to, but not identical with, the more classical notion of equidistribution involving an *infinite* sequence  $g : \mathbb{Z} \rightarrow G$ , in which one takes a limit as  $N \rightarrow \infty$ ; we refer to this latter concept as *asymptotic equidistribution* in order to distinguish it from the “single-scale” equidistribution considered here, in which one is working with a fixed (but unbounded)  $N$ . While there is a close analogy between the theory of asymptotic equidistribution and single-scale equidistribution, there does not seem to be a soft way to automatically transfer results from the former to the latter. Single-scale equidistribution is in fact much closer to the notion of  $\delta$ -*equidistribution* studied for instance in [24]; we refer readers to that paper for further discussion of the distinction between the different types of equidistribution.

*Example.* We consider the case when  $G/\Gamma = \mathbb{T}^d$  is a torus. Weyl’s equidistribution criterion, in our notation, then asserts that a limit map  $\mathcal{O} : [N]^k \rightarrow \mathbb{T}^d$  is equidistributed if and only if one has

$$\mathbb{E}_{n \in [N]^k} e(\xi \cdot \mathcal{O}(n)) = o(1)$$

for all standard  $\xi \in \mathbb{Z}^d \setminus \{0\}$ . One can also show (using some Fourier analysis) that  $\mathcal{O}$  will be totally equidistributed if and only if

$$\mathbb{E}_{n \in [N]^k} e(\xi \cdot \mathcal{O}(n)) e(\eta \cdot n) = o(1)$$

for all standard  $\xi \in \mathbb{Z}^d \setminus \{0\}$  and  $\eta \in \mathbb{Z}^k$ . As a consequence of this and some further Fourier analysis, we see that a one-dimensional linear orbit  $\mathcal{O} : [N] \rightarrow \mathbb{T}^d$  defined by  $\mathcal{O}(n) := \alpha n + \beta$  for some  $\alpha, \beta \in \mathbb{T}^d$  will be equidistributed or totally equidistributed in  $\mathbb{T}^d$  if and only if  $\alpha$  is not of the form  $q + O(N^{-1}) \bmod 1$  for some standard rational  $q \in \mathbb{Q}$ .

Given a standard filtered nilmanifold  $G/\Gamma$ , a *horizontal character* is a continuous standard homomorphism  $\xi : G \rightarrow \mathbb{T}$  which vanishes on  $\Gamma$ . We say that the character is *non-trivial* if it is not identically zero.

We have the following basic equidistribution criterion, generalising the torus example above.

**Theorem D.2** (Leibman theorem). *Let  $k \in \mathbb{N}^+$ , let  $N$  be an unbounded natural number, let  $G/\Gamma$  be an  $\mathbb{N}$ -filtered nilmanifold, and let  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow (G/\Gamma)_{\mathbb{N}})$  be a  $k$ -dimensional polynomial orbit, where  $\mathbb{Z}^k$  is given the degree filtration. Then on  $[N]^k$ , the following statements are equivalent:*

- (i)  $\mathcal{O}$  is totally equidistributed in the nilmanifold  $G/\Gamma$ ;
- (ii)  $\mathcal{O}$  is equidistributed in the nilmanifold  $G/\Gamma$ ;
- (iii)  $\mathcal{O}$  is equidistributed in the torus  $G/([G, G]\Gamma)$ , and
- (iv) There does not exist any non-trivial horizontal character  $\xi$  such that  $\xi \circ g$  is Lipschitz with constant  $O(1/N)$ .

*Proof.* See [24, Theorems 1.19, 2.9, 8.6] (where in fact a more quantitative strengthening of this equivalence is established). The analogue of this result for asymptotic equidistribution was established previously in [46] (and the result is classical in the case of linear sequences). The main difficulty is to show that (iv) implies (ii), which is the main content of [24, Theorem 2.9], which relies primarily on a certain van der Corput type equidistribution lemma for nilmanifolds.  $\square$

Theorem D.2 implies the following weak factorisation theorem.

**Theorem D.3** (Weak factorisation theorem). *Let  $k \in \mathbb{N}^+$ , let  $N$  be an unbounded natural number, let  $G/\Gamma$  be an  $\mathbb{N}$ -filtered nilmanifold and let  $g \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$ . Suppose that  $g$  is not totally equidistributed on  $[N]$  in  $G/\Gamma$ . Then one can factorise  $g = \varepsilon g' \gamma$ , where  $\varepsilon, g', \gamma \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$  have the following properties:*

- $\varepsilon$  is a bounded sequence on  $[N]^k$  with the  $i^{\text{th}}$  Taylor coefficient of size  $O(N^{-|i|})$  for each  $i \in \mathbb{N}^k$ ;
- $g'$  takes values in a standard proper rational subgroup  $G'$  of  $G$  (i.e.  $G'$  is a connected proper Lie subgroup of  $G$ , and  $\Gamma' := G' \cap \Gamma$  is cocompact in  $G$ ).
- $\gamma$  is periodic modulo  $\Gamma$  with a standard period  $q \in \mathbb{N}^+$ , thus  $\gamma(n + qv) = \gamma(n) \bmod \Gamma$  for all  $n, v \in {}^*\mathbb{Z}^k$ . Furthermore,  $\gamma$  takes values in a standard subgroup  $\tilde{\Gamma}$  of  $G$  which contains  $\Gamma$  as a subgroup.

*Proof.* See [24, Proposition 9.2]. The basic idea is to use the non-trivial horizontal character  $\xi$  generated by Theorem D.2 to cut out the subgroup  $G'$ . In order to keep  $G'$  connected, one needs to first factorise  $\xi = m\xi'$  where  $m$  is a standard positive integer and  $\xi'$  is an irreducible horizontal nilcharacter; this integer  $m$  is responsible for the periodic term  $\gamma$ .  $\square$

One can iterate this to obtain a “Ratner-type” theorem.

**Theorem D.4** (Factorisation theorem). *Let  $k \in \mathbb{N}^+$ , let  $N$  be an unbounded natural number, let  $G/\Gamma$  be a (filtered) nilmanifold, and let  $g \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$ . Then there exists a standard rational subgroup  $G'$  of  $G$  (i.e.  $G'$  is connected and  $G' \cap \Gamma$  is cocompact in  $G$ ) and a factorisation*

$$g(n) = \varepsilon(n)g'(n)\gamma(n)$$

where  $\varepsilon, g', \gamma \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$  have the following additional properties:

- $\varepsilon$  is a bounded sequence with the  $i^{\text{th}}$  Taylor coefficient of size  $O(N^{-|i|})$  for each  $i \in \mathbb{N}^k$ , and has Lipschitz constant  $O(1/N)$ ;
- $g'$  takes values in a standard proper rational subgroup  $G'$  of  $G$ , and is totally equidistributed in  $G'/\Gamma'$  whenever  $\Gamma'$  is any standard subgroup of  $G' \cap \Gamma$  of finite (standard) index.
- $\gamma$  is periodic modulo  $\Gamma$  with a standard period, and takes values in a standard discrete subgroup  $\tilde{\Gamma}$  of  $G$  which contains  $\Gamma$ .

This theorem is a close relative of [24, Theorem 1.19], and can be proven by the same methods; for the convenience of the reader we sketch a proof here.

*Proof.* Let us say that  $g$  can be represented using a standard rational subgroup  $G'$  of  $G$  if one has a factorisation  $g = \varepsilon g' \gamma$  which obeys all the conclusions of the theorem except for the total equidistribution of  $g'$ . Clearly,  $g$  can be represented using  $G$  itself, by setting  $\varepsilon$  and  $\gamma$  to be the identity and  $g' := g$ . By the principle of infinite descent<sup>12</sup> (using the fact that  $G$  has a finite standard dimension), we may thus find a standard rational subgroup  $G'$  which represents  $g$ , and is *minimal*

<sup>12</sup>The ability to use this principle is an advantage of the ultralimit setting. In the finitary setting, in which one needs to quantify such concepts as total equidistribution, periodicity, etc., one has to instead perform an iterative “dimension reduction argument” which requires one to manage many more parameters; see [24] for an example of this. See also the beginning of §10 for a related discussion.

in the sense that no proper standard rational subgroup of  $G'$  represents  $g$ . Let  $g = \varepsilon g' \gamma$  be the associated factorisation. It then suffices to show that  $g'$  is totally equidistributed in  $G'/\Gamma'$  for every standard finite index subgroup  $\Gamma'$  of  $\Gamma \cap G'$ .

Suppose for contradiction that this is not the case. Applying Theorem D.3, one can factorise  $g' = \varepsilon'' g'' \gamma''$  where  $\varepsilon''$  is a bounded sequence with Lipschitz constant  $O(1/N)$ ,  $\gamma''$  is periodic with a standard period, and takes values in a standard discrete subgroup  $\tilde{\Gamma}'$  that contains  $\Gamma'$ , and  $g''$  takes values in a proper rational subgroup  $G''$  of  $G'$ . One can enlarge  $\tilde{\Gamma}'$  to contain  $\Gamma$ , and this is easily verified to still be discrete. One can then show that the factorisation  $g = (\varepsilon \varepsilon'') g'' (\gamma'' \gamma)$  is a representation of  $g$  using  $G''$  (see [24, §10] for details), contradicting the minimality of  $G''$ .  $\square$

It will be convenient to convert the factorisation in Theorem D.4 into a more convenient form, eliminating the periodic factor  $\gamma$  and the slowly varying factor  $\varepsilon$  by passing to subprogressions.

**Theorem D.5** (Factorisation theorem, II). *Let  $k \in \mathbb{N}^+$ , let  $N$  be an unbounded natural number and let  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow (G/\Gamma)_{\mathbb{N}})$ . Then one can partition  $[N]^k$  into a bounded number of products  $P = P_1 \times \dots \times P_k$  of dense arithmetic subprogressions of  $[N]$ , such that for each  $P$  one has a polynomial  $\varepsilon \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$  which is bounded with Lipschitz constant  $O(1/N)$  on  $P$  and with the  $i^{\text{th}}$  Taylor coefficient of size  $O(N^{-|i|})$  for each  $i$ , a standard rational subgroup  $G_P$  of  $G$ , and a polynomial sequence  $g_P \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow (G_P)_{\mathbb{N}})$  totally equidistributed on  $G_P/\Gamma_P$  where  $(G_P)_{\mathbb{N}} := (G_P \cap G_i)_{i \in \mathbb{N}}$  and  $\Gamma_P := G_P \cap \Gamma$ , such that*

$$\mathcal{O}(n) = \varepsilon_P(n) g_P(n) {}^*\Gamma$$

for all  $n \in P$ . Furthermore, for each  $i \in \mathbb{N}^k$ , the horizontal Taylor coefficients  $\text{Taylor}_i(g)$  and  $\text{Taylor}_i(g_P)$  differ by  $O(N^{-|i|})$ . Finally, for two different products  $P, P'$  of progressions in this partition of  $[N]^k$ , the sequences  $g_P$  and  $g_{P'}$  are conjugate, with  $g_{P'} = \gamma_{P',P}^{-1} g_P \gamma_{P,P'}$  for some  $\gamma_{P,P'} \in G$  which is rational in the sense that  $\gamma_{P',P}^m \in \Gamma$  for some bounded positive integer  $m$ .

*Proof.* Write  $\mathcal{O}(n) = g(n) {}^*\Gamma$  for some  $g \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}}^k \rightarrow G_{\mathbb{N}})$ . We apply Theorem D.4 to obtain a rational standard subgroup  $G'$  and a factorisation  $g = \varepsilon g' \gamma$  with the stated properties. The sequence  $\gamma$  is periodic with a standard period, so we may partition  $[N]^k$  into a bounded number of products  $P = P_1 \times \dots \times P_k$  of dense arithmetic subprogressions of  $[N]$  on which  $\gamma = \gamma_P$  is constant. As  $\Gamma$  is cocompact, we may thus find  $\gamma'_P \in \gamma_P \Gamma$  which is bounded, thus  $\gamma'_P = O(1)$ . Note that  $\gamma'_P$  lives in a discrete group  $\tilde{\Gamma}$  and is thus standard. Since  $\Gamma$  is cocompact, it has finite index in  $\tilde{\Gamma}$ , which implies that  $\gamma'_P$  is rational, or equivalently that  $\gamma'_P$  has rational coefficients with respect to a Mal'cev basis [48] of  $G/\Gamma$ .

For  $n \in P$ , we can write

$$\mathcal{O}(n) = \varepsilon(n) g'(n) \gamma_P \Gamma = \varepsilon(n) \gamma'_P g_P(n) {}^*\Gamma$$

where  $g_P(n) := (\gamma'_P)^{-1} g'(n) \gamma'_P$  is the conjugate of  $g'(n)$  by  $\gamma'_P$ . Note that this gives the claim about the conjugate nature of  $g_P$  and  $g_{P'}$ .

As  $\gamma'_P$  is rational, the conjugate  $\gamma'_P \Gamma (\gamma'_P)^{-1}$  intersects  $\Gamma$  in a subgroup  $\Gamma'$  of finite index, which then has the property that  $\Gamma' \gamma'_P \subset \gamma'_P \Gamma$ . From this, we see that the conjugation operation  $g \mapsto (\gamma'_P)^{-1} g \gamma'_P$  on  $G$  descends to a continuous projection of  $G/\Gamma'$  to  $G/\Gamma$ , which maps  $g(n) {}^*\Gamma'$  to  $g_P(n) {}^*\Gamma$ . Since  $g(n)$  is totally equidistributed

on  $G'/(G' \cap \Gamma')$  by construction, we conclude that  $g_P$  is totally equidistributed on  $G_P/(G_P \cap \Gamma)$ , where  $G_P := (\gamma'_P)^{-1}G'\gamma'_P$  is the conjugate of  $G'$ . Note that  $G_P$  is also a standard rational subgroup of  $G$ . If we now set  $\varepsilon_P := \varepsilon\gamma'_P$ , we obtain all the claims except for the one about horizontal Taylor coefficients. But from the remarks following Definition 9.6 and the factorisations  $g = \varepsilon g'\gamma$ ,  $g_P = (\gamma'_P)^{-1}g'\gamma_P$  we have

$$\text{Taylor}_i(g) = \text{Taylor}_i(\varepsilon)\text{Taylor}_i(g')\text{Taylor}_i(\gamma)$$

and

$$\text{Taylor}_i(g_P) = \text{Taylor}_i(g').$$

Since  $\gamma$  takes values in  $\Gamma$ ,  $\text{Taylor}_i(\gamma)$  vanishes. Finally, by construction we have  $\text{Taylor}_i(\varepsilon) = O(N^{-|i|})$ . The claim follows.  $\square$

We can now give a criterion for when an average of the form  $\mathbb{E}_{n \in [N]} F(\mathcal{O}(n))$  is large.

**Theorem D.6** (Ratner-type theorem). *Let  $G/\Gamma$  be  $\mathbb{N}$ -filtered nilmanifold of some degree  $d$ , let  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$  be a polynomial orbit, and let*

$$F \in \text{Lip}({}^*(G/\Gamma) \rightarrow \overline{\mathbb{C}^{\omega}})$$

*be such that*

$$|\mathbb{E}_{n \in [N]} F(\mathcal{O}(n))| \gg 1.$$

*Then one has*

$$|\int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x)| \gg 1$$

*for some bounded  $\varepsilon \in G$  and some rational subgroup  $G_P$  of  $G$ , with the property that*

$$\pi_{\text{Horiz}_i(G)}(G_P \cap G_i) \geq \Xi_i^{\perp}$$

*where the horizontal space  $\text{Horiz}_i(G)$  and the projection map  $\pi_{\text{Horiz}_i(G)} : G_i \rightarrow \text{Horiz}_i(G)$  was defined in Definition 9.6,*

$$\Xi_i^{\perp} := \{x \in \text{Horiz}_i(G) : \xi_i(x) = 0 \text{ for all } \xi_i \in \Xi_i\}$$

*and  $\Xi_i \leq \widehat{\text{Horiz}_i(G/\Gamma)}$  is the group of all (standard) continuous homomorphisms  $\xi_i : \text{Horiz}_i(G/\Gamma) \rightarrow \mathbb{T}$  such that*

$$\xi_i(\text{Taylor}_i(\mathcal{O})) = O(N^{-i}).$$

One could also generalise this theorem to multidimensional orbits, but we will not need to do so in this paper. We will motivate this theorem with some examples after the proof.

*Proof.* By taking components we may assume that  $F$  is scalar-valued. Write  $\mathcal{O}(n) = g(n)*\Gamma$  for some  $g \in {}^*\text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow G_{\mathbb{N}})$ . We partition  $[N]$  into dense arithmetic progressions  $P$  induced from the partition of  $[N]$  coming from Theorem D.5 (using the Chinese remainder theorem and passing to dense subprogressions as necessary). By the pigeonhole principle, for at least one of these progressions  $P$  one has

$$|\mathbb{E}_{n \in P} F(g(n)*\Gamma)| \gg 1.$$

Now let  $\delta > 0$  be a small standard number to be chosen later. By further partitioning of  $P$  and the pigeonhole principle one can assume that  $P$  has diameter at most  $\delta N$  (note that the implied constant in the  $\gg$  notation remains independent of  $\delta$  when doing so). Then for any  $n_0 \in P$ ,  $\varepsilon_P(n)$  and  $\varepsilon_P(n_0)$  differ by  $O(\delta)$ , and so



(by the Lipschitz nature of  $F$ )  $F(g(n)^*\Gamma)$  differs from  $F(\varepsilon_P(n_0)g_P(n)^*\Gamma)$  by  $O(\delta)$ . Thus, for  $\delta$  sufficiently small, and setting  $\varepsilon := \varepsilon_P(n_0)$ , one has

$$|\mathbb{E}_{n \in P} F(\varepsilon g_P(n)^*\Gamma)| \gg 1.$$

Using the total equidistribution of  $g_P$ , we have

$$\mathbb{E}_{n \in P} F(\varepsilon g_P(n)^*\Gamma) = \int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x) + o(1)$$

and so

$$\int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x) \gg 1.$$

To finish the proof of Theorem D.6, we need to show that

$$\pi_{\text{Horiz}_i(G)}(G_P \cap G_i) \geq \Xi_i^\perp \quad (\text{D.4})$$

for all positive standard integers  $i$ , with  $\Xi_i$  as in Theorem D.6.

Fix  $i$ . To show the above claim, observe that  $g_P$  takes values in  $G_P$ , and so  $\text{Taylor}_i(g_P) \in \pi_{\text{Horiz}_i(G/\Gamma)}(G_P \cap G_i)$ . On the other hand,  $\text{Taylor}_i(g_P)$  differs from  $\text{Taylor}_i(g)$  by  $O(N^{-i})$ , and so

$$\text{dist}(\text{Taylor}_i(g), \pi_{\text{Horiz}_i(G/\Gamma)}(G_P \cap G_i)) = O(N^{-i}). \quad (\text{D.5})$$

Suppose the inclusion (D.4) failed. Then by duality (and the rational nature of  $G_P$ ), there exists a standard continuous homomorphism  $\xi_i : \text{Horiz}_i(G/\Gamma) \rightarrow \mathbb{T}$  outside of  $\Xi_i$  which annihilates  $\pi_i(G_P \cap G_i)$ . From (D.5), This implies that

$$\xi_i(\text{Taylor}_i(g)) = O(N^{-i}),$$

and thus  $\xi_i \in \Xi_i$  by definition of  $\Xi_i$ , contradiction. The claim follows.  $\square$

To get a feel for this proposition, let us first examine a simple special case, when  $G/\Gamma$  is just a two-dimensional torus  $\mathbb{T}^2$ , and  $\mathcal{O}$  is a linear orbit  $\mathcal{O}(n) := (\alpha n, \beta n)$  for some  $\alpha, \beta \in {}^*\mathbb{T}$ . We take  $F$  to be a standard Lipschitz function from  $\mathbb{T}^2$  to  $\mathbb{C}$ . Our hypothesis is then the assertion that

$$|\mathbb{E}_{n \in [N]} F(\alpha n, \beta n)| \gg 1.$$

The conclusion is then that

$$|\int_T F(\varepsilon + x) d\mu_T(x)| \gg 1$$

for some subtorus  $T := G_P/(G_P \cap \mathbb{Z}^2)$  of  $\mathbb{T}^2$ , where  $\varepsilon \in \mathbb{T}^2$  and  $G_P$  is a rational subgroup of  $\mathbb{R}^2$ . Furthermore,  $G_P$  contains the subgroup

$$\Xi_1^\perp := \{x \in \mathbb{R}^2 : \xi(x) = 0 \text{ for all } \xi \in \Xi_1\}$$

and  $\Xi_1$  is the subgroup of  $\mathbb{Z}^2$  defined by

$$\Xi_1 := \{\xi \in \mathbb{Z}^2 : \xi \cdot (\alpha, \beta) = O(N^{-1})\}.$$

We investigate some subcases of this result. First consider the case when  $\alpha, \beta$  are both within  $O(N^{-1})$  of standard rationals. Then  $\Xi_1$  is a finite index subgroup of  $\mathbb{Z}^2$ , and so  $\Xi_1^\perp$  is trivial. The conclusion is then simply the trivial conclusion that  $|F(\varepsilon)| \gg 1$  for some  $\varepsilon \in \mathbb{T}^2$ , which was of course obvious from the pigeonhole principle.

Now suppose that  $\beta$  is within  $O(N^{-1})$  of a standard rational  $p/q$  with  $p, q$  coprime, but that  $\alpha$  does not lie within  $O(N^{-1})$  of a standard rational. Then  $\Xi_1 = \{(0, qa) : a \in \mathbb{Z}\}$ , and so  $\Xi_1^\perp = \mathbb{R} \times \{0\}$ . The conclusion is now that

$|\int_{\mathbb{T}} F(x, \varepsilon) dx| \gg 1$  for some  $\varepsilon \in \mathbb{T}$ . This can also be seen directly by observing that on any subprogression of  $[N]$  of spacing  $q$  and length  $\delta N$  for some small  $\delta > 0$ , the orbit  $\mathcal{O}(n)$  is within  $O(\delta)$  of being equidistributed on a coset  $\mathbb{T} \times \{\varepsilon\}$  of  $\mathbb{T} \times \{0\}$  for some  $\varepsilon \in \mathbb{T}$ , with the implied constant in the  $O(\delta)$  notation independent of  $\delta$ . The claim then follows from the pigeonhole principle (choosing  $\delta$  sufficiently small, but still standard) and the Lipschitz nature of  $F$ .

Finally, suppose that  $\alpha, \beta$  are incommensurate in the sense that there does not exist any non-zero  $\xi \in \mathbb{Z}^2$  for which  $\xi \cdot (a, b) = O(N^{-1})$ . Then  $\Xi_1$  is trivial and so  $\Xi_1^\perp = \mathbb{R}^2$ . The claim is then that  $|\int_{\mathbb{T}^2} F(x, y) dx dy| \gg 1$ , which is also apparent from the equidistribution of  $\mathcal{O}$  in  $\mathbb{T}^2$  in this case.

One can also repeat the above example with the linear orbit  $n \mapsto (\alpha n, \beta n)$  replaced by a polynomial orbit such as  $n \mapsto (\alpha n^D, \beta n^D)$  for some standard  $D \geq 1$ . The discussion is identical, except that the  $O(N^{-1})$  errors must now be replaced by  $O(N^{-D})$ .

Now we consider the more general non-abelian setting, in which  $G/\Gamma$  is not necessarily a torus (i.e. we allow  $d$  to exceed 1). We first remark upon the “incommensurate”, “generic”, or “equidistributed” case when all the  $\Xi_i$  are trivial, i.e. there are no non-trivial relations of the form

$$\xi_i(\text{Taylor}_i(\mathcal{O})) = O(N^{-i}).$$

In this case,  $\Xi_i^\perp = \text{Horiz}_i(G)$  and so the maps  $\pi_i : G_P \cap G_i \rightarrow \text{Horiz}(G_i)$  are all surjective. This implies that all the horizontal spaces of the quotient group  $G/G_P$  are trivial, which one easily sees to imply that  $G/G_P$  itself must be trivial, i.e. that  $G_P = G$ . We conclude that  $|\int_{G/\Gamma} F d\mu| \gg 1$ . Indeed, in this case it turns out that  $\mathcal{O}$  is totally equidistributed and

$$\mathbb{E}_{n \in P_0} F(\mathcal{O}(n)) = \int_{G/\Gamma} F d\mu + o(1).$$

This fact can also be deduced from the *arithmetic counting lemma* [27, Theorem 1.11].

Finally, to illustrate how we actually use Theorem D.6 in practice, we consider a model problem in which we are given frequencies  $\alpha, \beta, \alpha', \beta' \in {}^*\mathbb{T}$  obeying the correlation property

$$|\mathbb{E}_{n \in [N]} e(\{\alpha n\}\beta n) e(\{\alpha' n\}\beta' n)| \gg 1, \quad (\text{D.6})$$

and we wish to conclude some constraints between these four frequencies; informally, the problem here is to determine for which frequencies  $\alpha, \beta, \alpha', \beta'$  can one have a non-trivial relationship between  $\{\alpha n\}\beta n$  and  $\{\alpha' n\}\beta' n$  (cf. (6.4)). Strictly speaking, for the analysis that we are about to give to apply, we must first replace the bracket polynomial expressions above by suitable vector-valued smoothings (or else develop analogues of the above equidistribution theory for piecewise Lipschitz nilsequences, as was done in the  $d = 2$  case in [28]), but to simplify the exposition we shall completely ignore this technical issue here.

Ignoring the technical issue alluded to above, we can express the left-hand side of (D.6) in the form  $|\mathbb{E}_{n \in [N]} F(\mathcal{O}(n))|$ , where  $G/\Gamma$  is the product Heisenberg nilmanifold of degree  $\leq 2$ , generated by four generators  $e_1, e_2, e'_1, e'_2$  with  $[e_1, e_2], [e'_1, e'_2]$  central (and with  $e_1, e_2$  commuting with  $e'_1, e'_2$ ),

$$\mathcal{O}(n) := e_2^{\beta n} e_1^{\alpha n} (e'_2)^{\beta' n} (e'_1)^{\alpha' n} \Gamma,$$

and  $F$  is a (piecewise) Lipschitz function on  $G/\Gamma$  obeying the vertical frequency property

$$F([e_1, e_2]^{t_{12}}[e'_1, e'_2]^{t'_{12}}x) = e(-t_{12} - t'_{12})F(x) \quad (\text{D.7})$$

for all  $x \in G/\Gamma$  and  $t_{12}, t'_{12} \in \mathbb{R}$ . Note that  $\text{Horiz}_1(G)$  is isomorphic to  $\mathbb{R}^4$  (being generated by the projections of  $e_1, e_2, e'_1, e'_2$  via  $\pi_{\text{Horiz}_1(G)}$ ), while  $\text{Horiz}_2(G)$  is trivial. Applying Theorem D.6, we conclude that

$$\left| \int_{G_P/\Gamma_P} F(\varepsilon x) d\mu(x) \right| \gg 1$$

for some bounded  $\varepsilon \in G$  and some rational subgroup  $G_P$  of  $G$ , with the property that

$$\pi_{\text{Horiz}_1(G)}(G_P) \geq \Xi_1^\perp \quad (\text{D.8})$$

where

$$\Xi_1 := \{\xi \in \mathbb{Z}^4 : \xi \cdot (\alpha, \beta, \alpha', \beta') = O(N^{-1})\}.$$

If the vertical group  $G_P \cap G_2$  contains any element  $[e_1, e_2]^{t_{12}}[e'_1, e'_2]^{t'_{12}}$  with  $-t_{12} - t'_{12} \neq 0$ , then from (D.7) we see that  $\int_{G_P} F(\varepsilon x) d\mu(x) = 0$ , a contradiction. We conclude that

$$G_P \cap G_2 \subset \langle [e_1, e_2][e'_1, e'_2]^{-1} \rangle_{\mathbb{R}}. \quad (\text{D.9})$$

This gives us some information concerning the group  $\Xi_1$ , and hence on the frequencies  $\alpha, \beta, \alpha', \beta'$ . Indeed, suppose that we are given two elements  $(a, b, a', b')$  and  $(c, d, c', d')$  in  $\Xi^\perp$ . By (D.8), we conclude that  $G_P$  contains two elements  $g, h$  such that

$$g = e_1^a e_2^b (e'_1)^{a'} (e'_2)^{b'} \bmod G_2$$

and

$$h = e_1^c e_2^d (e'_1)^{c'} (e'_2)^{d'} \bmod G_2.$$

Since  $g$  and  $h$  lie in  $G_P$ , the commutator

$$[g, h] = [e_1, e_2]^{ad-bc} [e'_1, e'_2]^{a'd'-b'c'}$$

must also lie in  $G_P$ . Comparing this with (D.9) we obtain an algebraic constraint on  $\Xi$  that prevents it from being too small, namely that

$$(ad - bc) + (a'd' - b'c') = 0 \quad (\text{D.10})$$

whenever  $(a, b, a', b'), (c, d, c', d') \in \mathbb{Z}^4$  are both orthogonal to  $\Xi$ ; thus the symplectic form (D.10) must vanish when restricted to  $\Xi^\perp$ .

For instance, suppose that  $(\alpha', \beta') = (\beta, \alpha)$ , but that  $\alpha, \beta$  are otherwise in general position (cf. (6.4)). Then  $\Xi$  is generated by  $(1, 0, 0, 1)$  and  $(0, 1, 1, 0)$ , so  $\Xi^\perp$  is generated by  $(1, 0, 0, -1)$  and  $(0, 1, -1, 0)$ , and one easily verifies the property. It is in principle possible to work out what other quadruples  $\alpha, \beta, \alpha', \beta'$  are permitted by Theorem D.6, but we will not compute this here.

## APPENDIX E. SOME BASIC PROPERTIES OF NILCHARACTERS AND SYMBOLS

In this appendix we establish some basic properties of nilcharacters and symbols; this material is broadly comparable to [28, §3].

Throughout this appendix,  $I$  is understood to be an ordering (see Definition 6.7).

We first begin with some basic closure properties of nilsequences.

**Lemma E.1** (Nilsequences are preserved by Lipschitz operations). *Let  $H$  be an  $I$ -filtered group, let  $J$  be a finite downset in  $I$ , and let  $\Omega$  be a limit subset of  ${}^*H$ . If  $\psi_i \in \text{Nil}^{\subset J}(\Omega \rightarrow \overline{\mathbb{C}}^{D_i})$  and  $D_i \in \mathbb{N}^+$  for  $i = 1, \dots, m$ , and  $F : \mathbb{C}^{D_1} \times \dots \times \mathbb{C}^{D_m} \rightarrow \mathbb{C}^D$  is a locally Lipschitz standard function, then  $F(\psi_1, \dots, \psi_m) \in \text{Nil}^{\subset J}(\Omega \rightarrow \mathbb{C}^D)$ .*

*Proof.* This follows immediately from Definition 6.19 and Example 6.13.  $\square$

As an immediate corollary we have the following.

**Corollary E.2** (Algebra property). *Let  $H$  be an  $I$ -filtered group, let  $J$  be a finite downset of  $I$ , and let  $\Omega$  be a limit subset of  $H$ . Then  $\text{Nil}^{\subset J}(\Omega \rightarrow \overline{\mathbb{C}})$  is a sub- $*$ -algebra of  $L^\infty(\Omega \rightarrow \overline{\mathbb{C}})$ , that is to say it is closed under pointwise multiplication, scalar multiplication by bounded constants, addition, and complex conjugation. Similarly,  $\text{Nil}^{\subset J}(\Omega \rightarrow \overline{\mathbb{C}}^\omega)$  is closed under complex conjugation, tensor product, and bounded linear combinations.*

*Remark.* From the example after Corollary B.5 we also see that if  $\psi \in L^\infty({}^*H \rightarrow \overline{\mathbb{C}}^\omega)$  is a nilsequence of degree  $\subset J$ , then so is any translate  $\psi(\cdot + h)$  or dilate  $\psi(q \cdot)$  of  $\psi$  for  $h \in {}^*H$  and  $q \in \mathbb{Z}$ .

**Lemma E.3** (Basic facts about nilcharacters). *Let  $H = (H, +)$  be an  $I$ -filtered abelian group for some  $I$ , let  $d \in I$ , and let  $\chi, \chi'$  be nilcharacters in  $\Xi^d({}^*H)$ . Then  $\chi \otimes \chi'$ ,  $\chi(\cdot + h)$ ,  $\chi(q \cdot)$ , and  $\overline{\chi}$  are also nilcharacters of degree  $\leq d$  for every  $h \in {}^*H$ , and  $q \in \mathbb{Z}$ .*

*More generally, if  $T : H' \rightarrow H$  is a (standard) filtered homomorphism from another  $I$ -filtered abelian group  $H' = (H', +)$  to  $H$ , then  $\chi \circ T$  is a nilcharacter in  $\Xi^d({}^*H')$ .*

*Finally, one has  $\Xi^{d'}({}^*H) \subset \Xi^d({}^*H')$  whenever  $d' < d$ .*

*Proof.* This follows from Corollary B.5 (cf. the example after that corollary, and Corollary E.2).  $\square$

From (6.5) it is trivial that a multidimensional polynomial of multidegree  $\subset J \cup J'$  can be decomposed as the sum of a multidimensional polynomial of multidegree  $\subset J$ , and a multidimensional polynomial of multidegree  $\subset J'$ . There is an analogous decomposition for nilcharacters.

**Lemma E.4** (Splitting lemma). *Let  $k \in \mathbb{N}^+$ , and let  $J, J'$  be finite downsets of  $\mathbb{N}^k$ . Let  $\psi \in \text{Nil}^{\subset J \cup J'}({}^*Z^k \rightarrow \overline{\mathbb{C}})$  be a nilsequence, and let  $\varepsilon > 0$  be standard. Then*

$$\|\psi(n) - \sum_{k=1}^K \psi_k(n) \psi'_k(n)\|_{L^\infty({}^*Z^k)} \leq \varepsilon$$

*where  $K$  is standard and for each  $1 \leq k \leq K$ ,  $\psi_k \in \text{Nil}^{\subset J}({}^*\mathbb{Z}^k \rightarrow \overline{\mathbb{C}})$  and  $\psi'_k \in \text{Nil}^{\subset J'}({}^*\mathbb{Z}^k \rightarrow \overline{\mathbb{C}})$ .*

*Proof.* We can write  $\psi = F \circ \mathcal{O}$ , where  $F \in \text{Lip}({}^*(G/\Gamma) \rightarrow \overline{\mathbb{C}})$ ,  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G/\Gamma)$ , and  $G/\Gamma$  is an  $\mathbb{Z}^k$ -filtered nilmanifold with degree  $\subset J \cup J'$ .

For each  $j \in J \cup J'$ , let  $e_{j,1}, \dots, e_{j,d_j}$  be a basis of generators for  $\Gamma_j$ . We may then lift  $G$  to the *universal nilpotent Lie group* that is formally generated by the  $e_{j,i}$ , subject to the constraint that any  $r-1$ -fold iterated commutator of the  $e_{j_1,i_1}, \dots, e_{j_r,i_r}$  with  $j_1 + \dots + j_r \notin J \cup J'$  vanishes, and similarly lift  $\Gamma$ ,  $F$ ,  $\mathcal{O}$  (using

Corollary B.10 for the latter). Thus we may assume without loss of generality that  $G$  is universal.

The degree  $\subset J \cup J'$  nilmanifold  $G/\Gamma$  projects down to the degree  $\subset J$  nilmanifold  $G/G_{>J}\Gamma$ , where  $G_{>J}$  is the group generated by the  $G_j$  for all  $j \in J' \setminus J$ . Similarly we have a projection from  $G/\Gamma$  to the degree  $\subset J'$  nilmanifold  $G/G_{>J'}\Gamma$ . The algebras  $\text{Lip}^*(G/G_{>J}\Gamma \rightarrow \overline{\mathbb{C}})$ ,  $\text{Lip}^*(G/G_{>J'}\Gamma \rightarrow \overline{\mathbb{C}})$  then pull back to subalgebras of  $\text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}})$ . By universality of  $G$ ,  $G_{>J}$  and  $G_{>J'}$  are disjoint. Thus, the union of these two algebras separate points in  $G/\Gamma$ . By the Stone-Weierstrass theorem, one can thus approximate  $F$  to arbitrary accuracy by products of elements from these algebras, and the claim follows.  $\square$

Next, we show that nilsequences can be decomposed into nilcharacters.

**Lemma E.5** (Fourier decomposition). *Let  $H$  be an  $I$ -filtered group, and let  $d \in I$ . If  $\psi \in \text{Nil}^{\leq d}(*H)$  and  $\varepsilon > 0$  is standard, then one can find a standard natural number  $m$ , and nilcharacters  $\chi_j \in \Xi^d(*H)$ , scalar nilsequences  $\psi_j \in \text{Nil}^{< d}(*H)$ , and bounded linear transformations  $T : \overline{\mathbb{C}}^{D_j} \rightarrow \overline{\mathbb{C}}^D$  for suitable dimensions  $D_j, D$  for each  $1 \leq j \leq m$  such that*

$$\|\psi - \sum_{j=1}^m T_j(\psi_j \otimes \chi_j)\|_{L^\infty(*H)} \leq \varepsilon.$$

*Proof.* It suffices to show this for scalar nilsequences  $\psi$ . Let  $G/\Gamma$  be an  $I$ -filtered nilmanifold of degree  $\leq d$ , let  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{\mathbb{C}})$ , and let  $\varepsilon > 0$ . We need to show that one can approximate  $F$  to uniform error at most  $\varepsilon$  by  $\sum_{j=1}^m T_j(F_j \otimes f_j)$ , where each  $F_j \in \text{Lip}(G/\Gamma \rightarrow S^{2D_j-1})$  has a vertical frequency,  $f_j \in \text{Lip}(G/\Gamma \rightarrow \mathbb{C})$  is invariant with respect to the  $G_d$  action (so that  $f_j$  descends to the quotient nilmanifold  $G/G_d\Gamma$ , which has degree  $< d$ ), and the  $T_j : \mathbb{C}^{D_j} \rightarrow \mathbb{C}$  are linear transformations.

Observe that the class of functions of the form  $\sum_{j=1}^m T_j(F_j \otimes f_j)$  form a complex algebra that are closed under conjugations. Thus by the Stone-Weierstrass theorem, it suffices to show functions of the form  $F \otimes f$ , where  $F \in \text{Lip}(G/\Gamma \rightarrow S^{2D-1})$  has a vertical frequency and  $f \in \text{Lip}(G/\Gamma \rightarrow \mathbb{C})$  and is invariant under  $G_d$ , separate points. This is trivial for two points which descend to distinct points on  $G/G_d\Gamma$ , so it suffices to do so for two points on a common  $G_d$  fibre. For this, it suffices by the definition of vertical frequency to show that for each  $g \in G_d$  with  $g \notin \Gamma_d$ , there exists a function  $F \in \text{Lip}(G/\Gamma \rightarrow S^{2D-1})$  has a vertical frequency  $\eta$  with  $\eta(g) \notin \mathbb{Z}$ .

The existence of a character  $\eta : G_d \rightarrow \mathbb{R}$  with  $\eta(g) \notin \mathbb{Z}$  is guaranteed by Pontryagin duality. Fixing such an  $\eta$ , we now perform the same construction used at the start of §6 (i.e. smoothly partition the base space  $G/G_d\Gamma$  into balls of small radius) to generate the desired function  $F$ .  $\square$

**Corollary E.6** (Correlation). *Let  $H$  be an  $I$ -filtered group, let  $d \in I$ , and let  $\Omega$  be a limit finite subset of  $*H$ . If  $f \in L^\infty(\Omega)$  is  $\leq d$ -biased, then  $f$  correlates with a nilcharacter in  $\Xi^d(\Omega)$ .*

*Proof.* We assume inductively that the claim has already been proven for all smaller values of  $d$ . We may assume that  $f$  is scalar. Applying Lemma E.5 for  $\varepsilon$  small enough, we see that  $f$  correlates with an expression of the form  $\sum_{j=1}^m T_j(\psi_j \otimes \chi_j)$ , and thus by the pigeonhole principle,  $f$  correlates with one of the  $\psi_j \otimes \chi_j$ , and thus  $f \overline{\chi_j}$  correlates with  $\psi_j$ . We can express the downset  $\{i \in I : i < d\}$  as the

finite union of downsets  $\{i \in I : i \leq d'\}$  for various  $d' < d$ . Applying Lemma E.4 repeatedly for sufficiently small  $\varepsilon$ , we thus see that  $f\overline{\chi_j}$  correlates with  $\prod_{d' \leq d} \psi_{d'}$ , where each  $\psi_{d'}$  is a nilsequence of degree  $\leq d'$ . Applying the inductive hypothesis repeatedly, we thus see that  $f\overline{\chi_j}$  correlates with  $\bigotimes_{d' < d} \chi_{d'}$  for some nilcharacters  $\chi_{d'}$  of degree  $\leq d'$ , and so  $f$  correlates with  $\chi_j \otimes \bigotimes_{d' < d} \chi_{d'}$ . The claim now follows from Lemma E.3.  $\square$

We turn now to a discussion of the basic properties of symbols. We begin by clearing up a small issue left over from §6: that of proving that the notion of equivalence we introduced in Definition 6.22 is indeed an equivalence relation. Recall that nilcharacters  $\chi$  and  $\chi'$  were said to be equivalent if  $\chi \otimes \overline{\chi'}$  is a nilsequence of degree strictly less than  $d$ .

**Lemma E.7.** *Equivalence of nilcharacters, thus defined, is an equivalence relation.*

*Proof.* The symmetry is obvious. For transitivity, suppose that  $\chi_1 \sim \chi_2$  and that  $\chi_2 \sim \chi_3$ . Then each component of

$$(\chi_1 \otimes \overline{\chi_2}) \otimes (\chi_2 \otimes \overline{\chi_3}) = \chi_1 \otimes (\overline{\chi_2} \otimes \chi_2) \otimes \overline{\chi_3}$$

is a nilsequence of degree strictly less than  $d$ . However the trace of  $\overline{\chi_2} \otimes \chi_2$  is 1, and so  $\chi_1 \otimes \overline{\chi_3}$  is a combination of the components of  $\chi_1 \otimes (\overline{\chi_2} \otimes \chi_2) \otimes \overline{\chi_3}$ . In particular, it is a nilsequence of degree strictly less than  $d$ .

To show reflexivity, we must confirm that  $\chi \otimes \overline{\chi}$  is a nilsequence of degree  $\prec d$  for any nilcharacter  $\chi \in \Xi^d(\Omega)$ . If we write  $\chi(n) = F(g(n)^*\Gamma)$ , where  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{S^\omega})$  has a vertical frequency  $\eta$ , we have

$$\chi \otimes \overline{\chi}(n) = (F \otimes \overline{F})(g(n)^*\Gamma).$$

Noting that  $F \otimes \overline{F}$  is invariant with respect to the  $G_d$  action, we may quotient out by this central group and represent  $\chi \otimes \overline{\chi}$  using a nilmanifold of degree  $\prec d$ .  $\square$

The space  $\text{Symb}^d(\Omega)$  has many nice properties.

**Lemma E.8** (Symbol calculus). *Let  $H = (H, +)$  be an abelian  $I$ -filtered group, let  $d \in I$ , and let  $\Omega$  be a limit subset of  ${}^*H$ .*

- (i) *If  $\chi, \chi' \in \Xi^d(\Omega)$  and  $\psi \in \text{Nil}^{<d}(\Omega)$ , and the components of  $\chi$  are bounded linear combinations of those of  $\chi' \otimes \psi$ , then  $\chi, \chi'$  are equivalent on  $\Omega$  and thus  $[\chi]_{\text{Symb}^d(\Omega)} = [\chi']_{\text{Symb}^d(\Omega)}$ .*
- (ii) *Conversely, if  $\chi, \chi' \in \Xi^d(\Omega)$  are equivalent, then  $\chi$  is a bounded linear combination of  $\chi' \otimes \psi$  for some  $\psi \in \text{Nil}^{<d}(\Omega)$ .*
- (iii)  *$\text{Symb}^d(\Omega)$  is an abelian group with the group operation induced from tensor product.*
- (iv) *If  $\chi \in \Xi^d({}^*H)$  and  $h \in {}^*H_i$  for some  $i > 0$ , then  $\chi$  and  $\chi(\cdot + h)$  are equivalent on  ${}^*H$  (and thus on  $\Omega$  also). Thus,  $[\chi(\cdot + h)]_{\text{Symb}^d(\Omega)} = [\chi]_{\text{Symb}^d(\Omega)}$ .*
- (v) *If  $H = \mathbb{Z}^k$  with either the multidegree or degree filtration,  $\chi \in \Xi^d({}^*H)$  and  $q \in \mathbb{Z}$ , then  $\chi^{\otimes q^{|d|}}$  and  $\chi(q \cdot)$  are equivalent on  ${}^*H$  (and thus on  $\Omega$  also), thus  $[\chi(q \cdot)]_{\text{Symb}^d(\Omega)} = q^{|d|}[\chi]_{\text{Symb}^d(\Omega)}$ .*
- (vi) (Pullback) *If  $T : {}^*\mathbb{Z}^k \rightarrow {}^*\mathbb{Z}^{k'}$  is a linear transformation, and  $\chi$  is a nilcharacter of degree  $d$  on  ${}^*\mathbb{Z}^{k'}$ , then  $\chi \circ T$  is a nilcharacter of degree  $d$  on  $\mathbb{Z}^k$ . Moreover, if  $\chi'$  is another nilcharacter of degree  $d$  on  ${}^*\mathbb{Z}^{k'}$  with  $[\chi]_{\text{Symb}^d({}^*\mathbb{Z}^{k'})} = [\chi']_{\text{Symb}^d({}^*\mathbb{Z}^{k'})}$ , then  $[\chi \circ T]_{\text{Symb}^d(\mathbb{Z}^k)} = [\chi' \circ T]_{\text{Symb}^d(\mathbb{Z}^k)}$ .*

- (vii) (Divisibility) *If  $H = \mathbb{Z}^k$  with either the multidegree or degree filtration,  $d \neq 0$ ,  $\chi \in \Xi^d(\Omega)$  and  $q \in \mathbb{N}^+$ , then there exists  $\tilde{\chi}$  such that  $[\chi]_{\text{Symb}^d(\Omega)} = q[\tilde{\chi}]_{\text{Symb}^d(\Omega)}$ .*

*Proof.* The claim (i) follows from the same argument used to prove reflexivity in Lemma E.7. For (ii), we proceed much as in the proof of transitivity in Lemma E.7: write  $\phi := \chi \otimes \overline{\chi'}$  and consider  $\chi' \otimes \phi = (\chi' \otimes \overline{\chi'}) \otimes \chi$ . Since 1 may be written as a linear combination of the components of  $\chi' \otimes \overline{\chi'}$ , the claim follows.

The claim (iii) follows easily from (i) and (ii). Part (iv) is more substantial. It should be compared to some of the consequences of the “bracket quadratic identities” developed in [28, Lemma 5.5].

From Definition 6.22, it suffices to show that the derivative  $\Delta_h \chi(n) := \chi(n+h) \otimes \overline{\chi(n)}$  lies in  $\text{Nil}^{<d}(*H)$ . We write  $\chi(n) = F(g(n)^* \Gamma)$ , where  $G/\Gamma$  is an  $I$ -filtered nilmanifold of degree  $\leq d$ ,  $g \in {}^* \text{poly}(H_I \rightarrow G_I)$ , and  $F \in \text{Lip}(* (G/\Gamma))$  has a vertical frequency  $\eta : G_d \rightarrow \mathbb{R}$ , then we have

$$\Delta_h \chi(n) = F((\partial_h g(n))g(n)^* \Gamma) \otimes \overline{F(g(n)^* \Gamma)}.$$

As  $g \in {}^* \text{poly}(H \rightarrow G)$  and  $h \in {}^* H_i$ , we have  $\partial_h g \in {}^* \text{poly}(H_I \rightarrow G_I^{+i})$ , where  $G_I^{+i} = (G_{j+i})_{j \in I}$  is the shifted filtration.

We now give  $G^2$  an  $I$ -filtration by defining  $(G^2)_j$  to be the group generated by  $G_{j+i} \times \text{id}$  and by the diagonal group  $\{(g, g) : g \in G_j\}$ . One easily verifies that this is a filtration, which is rational with respect to  $\Gamma^2$ . In particular, if we set  $G^\square := (G^2)_0$  and  $\Gamma^\square := \Gamma^2 \cap G^\square$ , we have that  $G^\square/\Gamma^\square$  is an  $I$ -filtered nilmanifold of degree  $\leq d$ . Furthermore, from Corollary B.4 we see that the map

$$\mathcal{O} : n \mapsto (\partial_h g(n)g(n), g(n))^* \Gamma^\square$$

lies in  ${}^* \text{poly}(H_I \rightarrow G_I^\square/\Gamma_I^\square)$ . We can thus write  $\Delta_h \chi = \tilde{F} \circ \mathcal{O}$ , where  $\tilde{F} \in \text{Lip}(* (G^\square/\Gamma^\square))$  is the function

$$\tilde{F}(x, y) := F(x) \otimes \overline{F(y)}.$$

This is still a degree  $\leq d$  representation. But observe from the vertical character nature of  $F$  that  $\tilde{F}$  is invariant with respect to the action of the group  $G_d^\square = \{(g, g) : g \in G_d\}$ . Thus we may quotient by this map and descend to a degree  $< d$  nilmanifold, and the claim follows.

Now we turn to (v), which is a similar claim to (iv). Writing  $\chi = F(g(n)^* \Gamma)$  as before, we reduce to showing that

$$n \mapsto F(g(qn)^* \Gamma) \otimes \overline{F(g(n)^* \Gamma)}^{\otimes q^{|d|}} \quad (\text{E.1})$$

can be represented as a nilsequence of degree  $< d$ , with the convention that  $F^{\otimes -q} = \overline{F}^{\otimes q}$  to deal with the case of negative exponents.

We give  $G^2$  a  $\mathbb{N}^k$ -filtration by declaring  $G_i^2$  to be the group generated by  $G_j \times G_j$  for all  $j > i$ , together with the set  $\{(g^{q^{|i|}}, g) : g \in G_i\}$ . From the Baker-Campbell-Hausdorff formula one easily sees that this is a filtration, which is rational with respect to  $\Gamma^2$ ; and so  $G_0^2/\Gamma_0^2$  is a degree  $\leq d$  nilmanifold. Also, from Taylor expansion (Lemma B.9) and Corollary B.4 we see that the map

$$\mathcal{O} : n \mapsto (g(qn), g(n))^* \Gamma_0^2$$

lies in  ${}^*\text{poly}(H \rightarrow G_0^2/\Gamma_0^2)$ . We then write (E.1) as  $n \mapsto \tilde{F}(\mathcal{O}(n))$ , where  $\tilde{F} \in \text{Lip}({}^*(G_0^2/\Gamma_0^2))$  is the function

$$\tilde{F}(x, y) := F(x) \otimes \overline{F}^{\otimes q^{|d|}}.$$

From the vertical character nature of  $F$ , we see that  $\tilde{F}$  is invariant with the action of  $G_2^2 = \{(g^{q^{|d|}}, g) : g \in G_d\}$ . Quotienting out by this group as in the proof of (iv) we obtain the claim.

The claim (vi) follows easily from Corollary B.5, so we now turn to (vii). We will prove this for the multidegree filtration, as the degree filtration is similar. As usual we write  $\chi = F(g(n)^*\Gamma)$ . Applying Taylor expansion (Lemma B.9) and the Baker-Campbell-Hausdorff formula, we may factorise

$$g(n) = \prod_{j \leq d} g_j^{n_j}$$

for some  $g_j \in G_j$ , where the product is over all multiindices  $j \leq d$  (arranged in some arbitrary fashion). Taking roots of each of the  $g_j$ , we may write  $g_j = (g'_j)^{q^{|j|}}$  for each  $j$ . We then have  $g(n) = g'(qn)$ , where  $g'$  is the polynomial sequence

$$g'(n) := \prod_{j \leq d} (g'_j)^{n_j}.$$

If we write  $\chi'(n) := F(g'(n)^*\Gamma)$ , we thus see that  $\chi' \in \Xi^d(\Omega)$  and  $\chi(n) = \chi'(qn)$ , so by (v),  $[\chi]_{\text{Symb}^d(\Omega)} = q^{|d|}[\chi']_{\text{Symb}^d(\Omega)}$ . The claim now follows by setting  $\tilde{\chi} := (\chi')^{\otimes q^{|d|-1}}$ .  $\square$

If  $P(n) = \alpha_0 + \dots + \alpha_d n^d$  is a polynomial of one variable  $n$  of degree  $d$ , then  $P$  is equal (up to degree  $< d$  errors) to the multilinear form  $Q(n, \dots, n)$ , where  $Q(n_1, \dots, n_d) := \alpha_d n_1 \dots n_d$ . A bit more generally, if  $P(n_1, \dots, n_k)$  is a polynomial of  $k$  variables  $n_1, \dots, n_k$  of multidegree  $d = (d_1, \dots, d_k)$ , then  $P$  is equal (up to degree  $< d$  errors) to a degree  $(1, \dots, 1)$  form  $Q(n_1, \dots, n_1, \dots, n_k, \dots, n_k)$ , where 1 is repeated  $|d|$  times and each  $n_i$  is repeated  $d_i$  terms. We may generalise this observaton to nilcharacters. We begin with the simpler  $k = 1$  case.

**Proposition E.9** (Multilinearisation,  $k = 1$  case). *Let  $d \in \mathbb{N}$  and  $\chi \in \Xi^d({}^*\mathbb{Z})$ . Then there exists  $\tilde{\chi} \in \Xi^{(1, \dots, 1)}({}^*\mathbb{Z}^d)$  (where 1 is repeated  $d$  times) such that the nilcharacter*

$$\chi' : n \mapsto \tilde{\chi}(n, \dots, n)$$

(where  $n$  is repeated  $d$  times) is equivalent to  $\chi$  in  $\Xi^d({}^*\mathbb{Z})$  (thus  $[\chi]_{\Xi^d({}^*\mathbb{Z})} = [\chi']_{\Xi^d({}^*\mathbb{Z})}$ ). Furthermore, one can select  $\tilde{\chi}(n_1, \dots, n_d)$  to be symmetric with respect to permutations of  $n_1, \dots, n_d$ .

To motivate this proposition, we present an “almost-example” of this proposition in action: if  $d = 2$  and  $\chi$  is the degree 2 almost-nilcharacter

$$\chi(n) := e(\{\alpha n\}\beta n),$$

(where the “almost” arises because the relevant function  $F$  is only piecewise Lipschitz rather than Lipschitz, as discussed at the start of §6) then one can take

$$\tilde{\chi}(n_1, n_2) := e\left(\frac{1}{2}\{\alpha n_1\}\beta n_2 + \frac{1}{2}\{\alpha n_2\}\beta n_1\right) \quad (\text{E.2})$$



which is a multidegree  $(1, 1)$  almost-nilcharacter, with  $\tilde{\chi}(n, n)$  equivalent (and in fact exactly equal, in this case) to  $\chi(n)$ . More generally, if we are able to represent a nilcharacter in terms of bracket polynomials of the correct degree and rank, then the above proposition becomes obvious by inspection. Such a representation is in fact possible (by extending the theory in [46]), but we will proceed here instead by using abstract algebraic constructions.

*Proof.* This will be a more complicated version of the argument used to establish claims (iv), (v) of Lemma E.8. It will be convenient for technical reasons to construct  $\tilde{\chi}$  so that  $\chi'$  is equivalent to  $\chi^{\otimes d!}$  rather than to  $\chi$  itself; to recover the original claim in the proposition, one simply appeals to Lemma E.8(vii).

We have  $\chi(n) = F(g(n)^*\Gamma)$  for some degree  $d$  nilmanifold  $G/\Gamma$ , some polynomial sequence  $g \in \text{poly}(\mathbb{Z}_{\mathbb{N}} \rightarrow (G/\Gamma)_{\mathbb{N}})$ , and some  $F \in \text{Lip}^*(G/\Gamma \rightarrow \overline{S}^{\omega})$ , obeying the vertical frequency property

$$F(g_d x) = e(\eta(g_d))F(x)$$

for all  $x \in G/\Gamma$  and  $g_d \in G_d$ , where  $\eta : G_d \rightarrow \mathbb{R}$  is a continuous homomorphism that maps  $\Gamma_d$  to the integers.

We now build the various components  $\tilde{G}, \tilde{\eta}, \tilde{g}, \tilde{F}$  required to construct  $\tilde{\chi}$ . (A simple example of this construction will be given after the end of this proof.)

The first step is build the multidegree  $(1, \dots, 1)$  nilpotent group  $\tilde{G}$ . We will construct this group via its nilpotent Lie algebra  $\log \tilde{G}$ . As a (real) vector space, this Lie algebra will be given as a direct sum

$$\log \tilde{G} := \oplus_{J \subset \{1, \dots, d\}} \log G_{|J|}.$$

For each  $J \subset \{1, \dots, d\}$ , let  $\iota_J : \log G_{|J|} \rightarrow \log \tilde{G}$  be the vector space embedding indicated by this direct sum, thus every element of  $\log \tilde{G}$  can be uniquely expressed in the form  $\sum_{J \subset \{1, \dots, d\}} \iota_J(x_J)$  for some  $x_J \in \log G_{|J|}$ .

Next, we endow  $\log \tilde{G}$  with a Lie bracket structure by declaring

$$[\iota_J(x_J), \iota_K(y_K)] = 0$$

whenever  $J, K \subset \{1, \dots, d\}$  intersect and  $x_J \in \log G_{|J|}$ ,  $y_K \in \log G_{|K|}$ , and

$$[\iota_J(x_J), \iota_K(y_K)] = \iota_{J \cup K}([x_J, y_K])$$

whenever  $J, K \subset \{1, \dots, d\}$  are disjoint and  $x_J \in \log G_{|J|}$ ,  $y_K \in \log G_{|K|}$ . One easily verifies that this operation obeys the axioms of a Lie bracket (i.e. it is bilinear, antisymmetric, and obeys the Jacobi identity), and so  $\log \tilde{G}$  is a Lie algebra.

We now give  $\log \tilde{G}$  a multidegree filtration. For any  $(a_1, \dots, a_d) \in \mathbb{N}^d$ , let  $\log \tilde{G}_{(a_1, \dots, a_d)}$  be the sub-Lie-algebra of  $\log \tilde{G}$  generated by the  $\iota_J(x_J)$  for which  $1_J(j) \geq a_j$  for each  $j = 1, \dots, d$ , and  $x_J \in \log G_{|J|}$ . One easily verifies that this is a multidegree filtration of multidegree  $(1, \dots, 1)$ , and so one can exponentiate to create a multidegree-filtered Lie group  $\tilde{G}$  of multidegree  $(1, \dots, 1)$  also.

We define a lattice  $\tilde{\Gamma}$  in  $\tilde{G}$  to be the group generated by  $\exp(M! \iota_J(\log \gamma_j))$  for all  $J \subset \{1, \dots, d\}$  and  $\gamma_j \in \Gamma_{|J|}$ , where  $M$  is a fixed natural number (depending only on  $d$ ) which we will assume to be sufficiently large. From the Baker-Campbell-Hausdorff formula we see that this is indeed a lattice, and so  $\tilde{G}/\tilde{\Gamma}$  is a nilmanifold. For  $M$  large enough, we see from further application of the Baker-Campbell-Hausdorff formula that  $\tilde{\Gamma}_{(1, \dots, 1)}$  is contained in  $\iota_{(1, \dots, 1)}(\log \Gamma_d)$ .

Next, we define a vertical frequency  $\tilde{\eta}$  on  $\tilde{G}_{(1,\dots,1)}$  by setting

$$\tilde{\eta}(\iota_{(1,\dots,1)}(\log g_d)) := \eta(g_d).$$

One easily verifies that  $\tilde{\eta}$  is a vertical frequency (here we use the inclusion  $\tilde{\Gamma}_{(1,\dots,1)} \subset \iota_{(1,\dots,1)}(\log \Gamma_d)$  and the central nature of  $G_{(1,\dots,1)}$ ).

Now let  $\tilde{F} \in \text{Lip}^*(\tilde{G}/\tilde{\Gamma} \rightarrow \overline{S^\omega})$  be a function with vertical frequency  $\tilde{\eta}$ ; such a function can be constructed using partitions of unity as in (6.3).

The next step is to define  $\tilde{g}$ . From Lemma B.9 and many applications of the Baker-Campbell-Hausdorff formula, we may write

$$g(n) = \prod_{j=0}^d g_j^{n_j}$$

for some coefficients  $g_j \in G_j$ . We then write

$$\tilde{g}(n_1, \dots, n_d) := \prod_{j=0}^d \exp(j! \sum_{J \subset \{1,\dots,d\}: |J|=j} (\prod_{i \in J} n_i) \iota_J(\log g_j)).$$

Observe that each individual monomial

$$(n_1, \dots, n_d) \mapsto \exp(j! (\prod_{i \in J} n_i) \iota_J(\log g_j))$$

with  $0 \leq j \leq d$  and  $|J| = j$  is a polynomial map in  ${}^*\text{poly}(\mathbb{Z}_{\mathbb{N}^d}^d \rightarrow \tilde{G}_{\mathbb{N}^d})$ , so by Corollary B.4 and the Baker-Campbell-Hausdorff formula we see that the same is true for  $\tilde{g}$ .

Finally, we set

$$\tilde{\chi}(n_1, \dots, n_d) := \tilde{F}(\tilde{g}(n_1, \dots, n_d) * \tilde{\Gamma}).$$

By construction,  $\tilde{\chi} \in \Xi^{(1,\dots,1)}({}^*\mathbb{Z}^d)$ , which by Lemma E.8(vi) (and the embeddings in Example 6.11) implies that  $\chi' \in \Xi^d({}^*\mathbb{Z})$ . It is also clear that  $\tilde{\chi}$  is symmetric with respect to permutations of the  $n_1, \dots, n_d$ . It remains to show that  $\chi'$  is equivalent to  $\chi^{\otimes d!}$  in  $\Xi^d({}^*\mathbb{Z})$ , or in other words that the sequence

$$n \mapsto \chi(n)^{\otimes d!} \otimes \overline{\tilde{\chi}}(n, \dots, n)$$

is a nilsequence of degree  $< d$ . We expand this sequence as

$$(F^{\otimes d!} \otimes \overline{\tilde{F}}) \left( \prod_{j=0}^d (g_j, \exp(j! \sum_{J \subset \{1,\dots,d\}: |J|=j} \iota_J(\log g_j)))^{n_j} * (\Gamma \times \tilde{\Gamma}) \right).$$

The function  $F \otimes \overline{\tilde{F}}$  is a Lipschitz function on the nilmanifold  $(G \times \tilde{G})/(\Gamma \times \tilde{\Gamma})$ . Let  $G^*$  be the subgroup of  $G \times \tilde{G}$  defined as

$$G^* := \{(g_d, \exp(d! \iota_{(1,\dots,1)}(\log g_d)) : g_d \in G_d\} \leq G_d \times \tilde{G}_{(1,\dots,1)}.$$

This is a rational central subgroup. As  $F$  and  $\tilde{F}$  have vertical frequencies  $\eta$  and  $\tilde{\eta}$  respectively, we see that  $F \otimes \overline{\tilde{F}}$  is invariant in the  $G^*$  direction, and thus descends to a Lipschitz function  $F'$  on the nilmanifold  $G'/\Gamma'$ , where  $G' := (G \times \tilde{G})/G^*$  and  $\Gamma'$  is the projection of  $\Gamma \times \tilde{\Gamma}$  to  $G'$ . We thus have

$$\chi(n) \otimes \overline{\tilde{\chi}}(n, \dots, n) = F' \left( \prod_{j=1}^d (g'_j)^{n_j} * \Gamma' \right) \quad (\text{E.3})$$

where  $g'_j$  is the projection of  $(g_j, \exp(j! \sum_{J \subset \{1, \dots, d\}: |J|=j} \iota_J(\log g_j)))$  to  $G'$ .

We now give  $G'$  a degree filtration by defining  $G'_j$  to be the group generated by elements of the form

$$(h_j, \exp(j! \sum_{J \subset \{1, \dots, d\}: |J|=j} \iota_J(\log h_j))) \bmod G^*$$

for  $h_j \in G_j$ , together with elements of the form

$$(h_{j+1}, \text{id}), (\text{id}, \exp(\iota_J(\log h_{j+1}))) \bmod G^*$$

for  $h_{j+1} \in G_{j+1}$  and  $J \subset \{1, \dots, d\}$  with  $|J| = j + 1$ . By a tedious number of applications of the Baker-Campbell-Hausdorff formula, we see that this is a filtration of degree  $< d$  (here we use the fact that every set of cardinality  $j + k$  has  $\frac{(j+k)!}{j!k!}$  partitions into a set  $J$  of cardinality  $j$  and a set  $K$  of cardinality  $k$ , which cancels the  $j!$  prefactors appearing in the definition of  $G'_j$ ). By construction,  $g'_j \in G'_j$ . Thus the right-hand side of (E.3) is a nilsequence of degree  $< d$ , and the claim follows.  $\square$

*Example.* We illustrate the above proposition with the simple  $d = 2$  example mentioned before the proof. We consider a nilcharacter  $\chi$  that is a vector-valued smoothing of the sequence  $n \mapsto e(\{\alpha n\}\beta n)$  for some fixed frequencies  $\alpha, \beta \in {}^*\mathbb{T}$ , which we will write schematically as

$$\chi(n) \sim e(\{\alpha n\}\beta n).$$

As discussed in §6, such a nilcharacter arises from the Heisenberg nilmanifold (6.1) with the polynomial sequence

$$g(n) = e_2^{\beta n} e_1^{\alpha n}$$

and vertical character  $\eta([e_1, e_2]^{t_{12}}) := -t_{12}$ . We may Taylor expand  $g$  as

$$g(n) = g_1^n g_2^{n^2}$$

where  $g_1 := \exp(\alpha \log e_1 + \beta \log e_2) = e_1^\alpha e_2^\beta [e_1, e_2]^{-\alpha\beta/2}$  and  $g_2 := [e_1, e_2]^{-\alpha\beta/2}$ .

The nilpotent Lie algebra  $\log \tilde{G}$  is the seven-dimensional vector space

$$\log \tilde{G} = \log G \oplus \log G \oplus \log G_{12}$$

with a basis of this space given by

$$\iota_1(\log e_1), \iota_1(\log e_2), \iota_1(\log[e_1, e_2]), \iota_2(\log e_1), \iota_2(\log e_2), \iota_2(\log[e_1, e_2]), \iota_{12}(\log[e_1, e_2]). \quad (\text{E.4})$$

The Lie algebra commutation relations on basis elements are given by the formulae

$$\begin{aligned} [\iota_1(\log e_1), \iota_2(\log e_2)] &= \iota_{12}(\log[e_1, e_2]) \\ [\iota_1(\log e_2), \iota_2(\log e_1)] &= -\iota_{12}(\log[e_1, e_2]) \end{aligned}$$

with all other pairs of basis elements commuting. This gives a nilpotent Lie group  $\tilde{G}$  generated (as a Lie group) by the exponentials of (E.4), which we will label as

$$\begin{aligned} a_1 &:= \exp(\iota_1(\log e_1)) \\ a_2 &:= \exp(\iota_1(\log e_2)) \\ a_{12} &:= \exp(\iota_1(\log e_{12})) \\ b_1 &:= \exp(\iota_2(\log e_1)) \\ b_2 &:= \exp(\iota_2(\log e_2)) \\ b_{12} &:= \exp(\iota_2(\log e_{12})) \\ c_{12} &:= \exp(\iota_{12}(\log e_{12})), \end{aligned}$$

thus one has the group commutation relations

$$[a_1, b_2] = c_{12}; \quad [a_2, b_1] = c_{12}^{-1}$$

with all other pairs of generators commuting. The generators  $a_{12}, b_{12}$  will play no essential role in the analysis that follows and may be ignored by the reader.

The group  $\tilde{G}$  is a multidegree  $(1, 1)$  filtered Lie group with filtration

$$\begin{aligned} \tilde{G}_{(0,0)} &:= \tilde{G}; \\ \tilde{G}_{(1,0)} &:= \langle a_1, a_2, a_{12}, c_{12} \rangle_{\mathbb{R}}; \\ \tilde{G}_{(0,1)} &:= \langle b_1, b_2, b_{12}, c_{12} \rangle_{\mathbb{R}}; \\ \tilde{G}_{(1,1)} &:= \langle c_{12} \rangle_{\mathbb{R}}. \end{aligned}$$

To construct  $\tilde{\Gamma}$ , we may take  $M = 1$ , so that

$$\tilde{\Gamma} := \langle a_1, a_2, a_{12}, b_1, b_2, b_{12}, c_{12} \rangle.$$

From the Baker-Campbell-Hausdorff formula one sees that

$$\tilde{\Gamma}_{(1,1)} := \tilde{\Gamma} \cap \tilde{G}_{(1,1)} = \langle c_{12} \rangle.$$

A typical element of  $\tilde{G}/\tilde{\Gamma}$  can be parameterised as

$$a_1^{r_1} a_2^{r_2} a_{12}^{r_{12}} b_1^{s_1} b_2^{s_2} b_{12}^{s_{12}} c_{12}^{t_{12}} \tilde{\Gamma}$$

for  $r_1, r_2, r_{12}, s_1, s_2, s_{12}, t_{12} \in I_0$ .

The polynomial sequence  $\tilde{g}$  is given as

$$\begin{aligned} \tilde{g}(n_1, n_2) &:= \exp(n_1 \iota_1(\log g_1) + n_2 \iota_2(\log g_1)) \exp(2n_1 n_2 \iota_{12}(\log g_2)) \\ &= \exp(\alpha n_1 \log a_1 + \beta n_1 \log a_2 + \alpha n_2 \log b_1 + \beta n_2 \log b_2) \times \\ &\quad \times \exp(-\alpha \beta n_1 n_2 \log c_{12}) \end{aligned}$$

which by the Baker-Campbell-Hausdorff formula expands to

$$\tilde{g}(n_1, n_2) = a_1^{\alpha n_1} a_2^{\beta n_1} b_1^{\alpha n_2} b_2^{\beta n_2} c_{12}^{-\alpha \beta n_1 n_2}.$$

This is clearly a polynomial sequence. If we then let  $\tilde{\eta} : \tilde{G}_{(1,1)} \rightarrow \mathbb{R}$  be the vertical character

$$\tilde{\eta}(\exp(t_{12} \iota_{12}(\log[e_1, e_2]))) := -t_{12}$$

and let  $\tilde{F} : \tilde{G}/\tilde{\Gamma} \rightarrow S^1$  be the (piecewise) Lipschitz function

$$\tilde{F}(a_1^{r_1} a_2^{r_2} a_{12}^{r_{12}} b_1^{s_1} b_2^{s_2} b_{12}^{s_{12}} c_{12}^{t_{12}} \tilde{\Gamma}) := e(-t_{12})$$

for  $r_1, r_2, r_{12}, s_1, s_2, s_{12}, t_{12} \in I_0$ , then the sequence

$$\tilde{\chi}(n_1, n_2) := \tilde{F}(\tilde{g}(n_1, n_2)^* \tilde{\Gamma})$$

is almost a nilcharacter of multidegree  $(1, 1)$ , if we make the usual cheat of ignoring the fact that  $\tilde{F}$  is only piecewise Lipschitz rather than Lipschitz.

Now let us look at the diagonal sequence

$$\tilde{\chi}(n, n) = \tilde{F}(a_1^{\alpha n} a_2^{\beta n} b_1^{\alpha n} b_2^{\beta n} c_{12}^{-\alpha \beta n^2} * \tilde{\Gamma}).$$

A brief computation using the Baker-Campbell-Hausdorff formula shows that one can rewrite

$$a_1^{\alpha n} a_2^{\beta n} b_1^{\alpha n} b_2^{\beta n} c_{12}^{-\alpha \beta n^2} * \tilde{\Gamma}$$

as

$$a_1^{\{\alpha n\}} a_2^{\{\beta n\}} b_1^{\{\alpha n\}} b_2^{\{\beta n\}} c_{12}^{(\alpha n - \{\alpha n\})(\beta n - \{\beta n\}) - (\beta n - \{\beta n\})(\alpha n - \{\alpha n\}) - \alpha \beta n^2} * \tilde{\Gamma}.$$

Noting that  $(\alpha n - \{\alpha n\})(\beta n - \{\beta n\})$  is an integer (cf. (6.4)), we can write the  $c_{12}$  exponent modulo 1 as

$$\{\alpha n\}\{\beta n\} - 2\{\alpha n\}\beta n \bmod 1$$

and thus

$$\tilde{\chi}(n, n) = e(2\{\alpha n\}\beta n) e(-\{\alpha n\}\{\beta n\}).$$

The second factor  $e(-\{\alpha n\}\{\beta n\})$  is a piecewise Lipschitz function of  $(\alpha n \bmod 1, \beta n \bmod 1)$  and is thus almost a 1-step nilsequence. We thus see that  $\tilde{\chi}(n, n)$  is almost equivalent (as a degree 2 almost nilcharacter) to  $\chi(n)^2$ . To eliminate the exponent of 2, one can go back to the start of the argument and replace  $\beta$  (for instance) by  $\beta/2$ . The reader may verify that once one does so, the almost nilcharacter  $\tilde{\chi}$  is essentially equal to (E.2).

Finally, we mention that with the above example, the group  $G^*$  takes the form

$$G^* := \{([e_1, e_2]^{t_{12}}, c_{12}^{2t_{12}}) : t_{12} \in \mathbb{R}\}$$

and the group  $G' := (G \times \tilde{G})/G^*$  has the degree 1 filtration

$$G'_0 := G'$$

$$G'_1 := \{(e_1^{t_1} e_2^{t_2} [e_1, e_2]^{t_{12}}, a_1^{t_1} a_2^{t_2} a_{12}^{t'_{12}} b_1^{t_1} b_2^{t_2} b_{12}^{t'_{12}} c_{12}^{t''_{12}} : t_1, t_2, t_{12}, t'_{12}, t''_{12} \in \mathbb{R}\} \bmod G^*.$$

One can verify by hand that this is indeed a degree 1 filtration on  $G'$ , which explains why  $\chi(n)^2 \tilde{\chi}(n, n)$  is a degree 1 (almost) nilsequence.

This concludes the discussion of the example. Now we generalise Proposition E.10 to higher  $k$ .

**Theorem E.10** (Multilinearisation). *Let  $\Omega$  be a limit subset of  $\mathbb{Z}^k$ , which we give the multidegree filtration. Let  $d = (d_1, \dots, d_k) \in \mathbb{N}^k$ , and  $\chi \in \Xi^d(\Omega)$ . Then there exists  $\tilde{\chi} \in \Xi^{(1, \dots, 1)}(*\mathbb{Z}^{|d|})$  (where 1 is repeated  $|d|$  times) such that the nilcharacter*

$$\chi' : (n_1, \dots, n_k) \mapsto \tilde{\chi}(n_1, \dots, n_1, n_2, \dots, n_2, \dots, n_k, \dots, n_k)$$

(where each  $n_i$  is repeated  $d_i$  times) is equivalent to  $\chi$  in  $\Xi^d(\Omega)$  (thus  $[\chi]_{\Xi^d(\Omega)} = [\chi']_{\Xi^d(\Omega)}$ ). Furthermore, one can select

$$\tilde{\chi}(n_{1,1}, \dots, n_{1,d_1}, n_{2,1}, \dots, n_{2,d_2}, \dots, n_{k,1}, \dots, n_{k,d_k})$$

to be symmetric with respect to the permutation of  $n_{i,1}, \dots, n_{i,d_i}$  for each  $i = 1, \dots, k$ .

*Proof.* Without loss of generality we may take  $\Omega = *\mathbb{Z}^k$ . The argument is exactly the same as that used to establish Proposition E.9 except that the notation is more complicated. Accordingly, we will focus primarily on the notational setup in this proof.

As before, it will suffice to make  $\chi'$  equivalent to  $\chi^{\otimes d!}$  rather than  $\chi$ , where  $d! := d_1! \dots d_k!$ . We have  $\chi(n) = F(g(n)*\Gamma)$  for some multidegree  $d$  nilmanifold  $G/\Gamma$ , some polynomial sequence  $g \in *\text{poly}(\mathbb{Z}_{\mathbb{N}^k}^k \rightarrow (G/\Gamma)_{\mathbb{N}^k})$ , and some  $F \in \text{Lip}(* (G/\Gamma) \rightarrow \overline{S^\omega})$ , obeying the vertical frequency property

$$F(g_d x) = e(\eta(g_s))F(x)$$

for all  $x \in G/\Gamma$  and  $g_d \in G_d$ , where  $\eta : G_d \rightarrow \mathbb{R}$  is a vertical frequency.

As before, we begin by building the nilpotent Lie algebra  $\log \tilde{G}$ . As a (real) vector space, this Lie algebra will be given as a direct sum

$$\log \tilde{G} := \oplus_{J \subset \{1, \dots, |d|\}} \log G_{\|J\|}$$

where  $\|J\| \in \mathbb{N}^k$  is the vector

$$\|J\| := (|J \cap \{d_1 + \dots + d_{i-1} + 1, \dots, d_1 + \dots + d_i\}|)_{1 \leq i \leq k}.$$

For each  $J \subset \{1, \dots, d\}$ , let  $\iota_J : \log G_{\|J\|} \rightarrow \log \tilde{G}$  be the vector space embedding indicated by this direct sum. Next, we endow  $\log \tilde{G}$  with a Lie bracket structure by declaring

$$[\iota_J(x_J), \iota_K(y_K)] = 0$$

whenever  $J, K \subset \{1, \dots, d\}$  intersect and  $x_J \in \log G_{\|J\|}$ ,  $y_K \in \log G_{\|K\|}$ , and

$$[\iota_J(x_J), \iota_K(y_K)] = \iota_{J \cup K}([x_J, y_K])$$

whenever  $J, K \subset \{1, \dots, d\}$  are disjoint and  $x_J \in \log G_{\|J\|}$ ,  $y_K \in \log G_{\|K\|}$ . As before, one easily verifies the Lie bracket axioms.

We now give  $\log \tilde{G}$  a multidegree filtration. For any  $(a_1, \dots, a_{|d|}) \in \mathbb{N}^{|d|}$ , let  $\log \tilde{G}_{(a_1, \dots, a_{|d|})}$  be the sub-Lie-algebra of  $\log \tilde{G}$  generated by the  $\iota_J(x_J)$  for which  $1_J(j) \geq a_j$  for each  $j = 1, \dots, |d|$ , and  $x_J \in G_{\|J\|}$ . As before, this is a multidegree filtration of multidegree  $(1, \dots, 1)$ , and exponentiates to create a multidegree-filtered Lie group  $\tilde{G}$  of multidegree  $(1, \dots, 1)$  also.

We define a lattice  $\tilde{\Gamma}$  in  $\tilde{G}$  to be the group generated by  $\exp(M! \iota_J(\log \gamma_j))$  for all  $J \subset \{1, \dots, |d|\}$  and  $\gamma_j \in \Gamma_{\|J\|}$ . Again, this creates a nilmanifold  $\tilde{G}/\tilde{\Gamma}$  is a nilmanifold, and for  $M$  large enough,  $\tilde{\Gamma}_{(1, \dots, 1)}$  is contained in  $\iota_{(1, \dots, 1)}(\log \Gamma_d)$ .

As before, we define a vertical frequency  $\tilde{\eta}$  on  $\tilde{G}_{(1, \dots, 1)}$  by the exact same formula:

$$\tilde{\eta}(\iota_{(1, \dots, 1)}(\log g_d)) := \eta(g_d),$$

and then construct  $\tilde{F} \in \text{Lip}(*(\tilde{G}/\tilde{\Gamma}) \rightarrow \overline{S^\omega})$  with vertical frequency  $\tilde{\eta}$ .

The next step is to define  $\tilde{g}$ . As before, we have the Taylor expansion

$$g(n) = \prod_{j \leq d} g_j^{n_j}$$

for some coefficients  $g_j \in G_j$ , where  $j = (j_1, \dots, j_k)$  now ranges over multi-indices less than or equal to  $d$ , arranged in some arbitrary order (e.g. lexicographical will suffice). We then write

$$\tilde{g}(n_1, \dots, n_{|d|}) := \prod_{j \leq d} \exp(j! \sum_{J \subset \{1, \dots, |d|\}: \|J\|=j} (\prod_{i \in J} n_i) \iota_J(\log g_j)),$$

recalling that  $j! := j_1! \dots j_k!$ . As before, one verifies that  $\tilde{g}$  is a polynomial map.

Finally, we set

$$\tilde{\chi}(n_1, \dots, n_{|d|}) := \tilde{F}(\tilde{g}(n_1, \dots, n_{|d|})^* \tilde{\Gamma}).$$

The rest of the argument proceeds exactly as in Proposition E.9, the main difference being that  $d$  is replaced with  $|d|$ , and  $|J|$  with  $\|J\|$ , whenever necessary; we omit the details.  $\square$

Now we show how nilcharacters interact with the concept of bias.

**Lemma E.11** (Bias lemma). *Let  $k, d \in \mathbb{N}^+$  with  $d \geq 2$ , let  $\chi$  be a degree  $d$  nilcharacter on  ${}^*\mathbb{Z}^k$  (with the degree filtration), and let  $N$  be an unbounded limit natural number. Let  $\Omega$  be a convex polytope in  $[[N]]^k$ , let  $P_1, \dots, P_k$  be dense subprogressions of  $[N]$ . Suppose that  $1_\Omega(n)1_{P_1 \times \dots \times P_k} \chi(n)$  is  $< d$ -biased on  $[[N]]^k$ . Then on  $[[N]]^k$ ,  $\chi$  is equal to a nilsequence of degree  $< d$ .*

*Remark.* Note that the claim fails for  $d = 1$ , even when  $k = 1$ ; if  $q > 1$  is a bounded integer, then the degree 1 nilcharacter  $n \mapsto e(n/q)$  is of course biased on progression of spacing  $q$ , but not on the original interval  $[[N]]$ . However, this is a purely “degree 1” obstruction and vanishes for higher degree.

*Proof.* Write  $P := P_1 \times \dots \times P_k$ . By Corollary E.6,  $1_\Omega(n)1_P \chi(n)$  correlates with a nilcharacter of degree  $d - 1$ ; we may absorb this nilcharacter into  $\chi$ , and assume that  $1_\Omega(n)1_P \chi(n)$  is in fact biased.

By partitioning  $\Omega \cap P$  into the product  $P' = P'_1 \times \dots \times P'_k$  of dense progressions of  $[[N]]$  (using [23, Corollary A.2] to control the error), we see that there exists such a product  $P' = P'_1 \times \dots \times P'_k$  for which

$$|\mathbb{E}_{n \in P'} \chi(n)| \gg 1$$

Write  $\chi = F \circ \mathcal{O}$  for some degree  $\leq d$  nilmanifold  $G/\Gamma$ , some  $F \in \text{Lip}^*(G/\Gamma)$  with a vertical frequency  $\eta$ , and some  $\mathcal{O} \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G/\Gamma)$ . Applying Theorem D.5 and using the pigeonhole principle to refine the progressions  $P'_1, \dots, P'_k$  if necessary, we may assume without loss of generality that we can factorise

$$\mathcal{O}(n) = \varepsilon_{P'}(n) g_{P'}(n)^* \Gamma$$

for all  $n \in P'$ , where  $g_{P'} \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G_{P'})$  is totally equidistributed on  $G_{P'}/\Gamma_{P'}$  for some standard rational subgroup  $G_{P'}$  of  $G$ , and  $\varepsilon_{P'} \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G)$  being bounded and having Lipschitz constant  $O(1/N)$  on  $P$ , and with the  $i^{\text{th}}$  Taylor coefficients of size  $O(N^{-|i|})$  for each  $i \in \mathbb{N}^k$ .

For any  $n, n_{P'} \in P'$ , we have from the Lipschitz nature of  $\varepsilon_{P'}$  that

$$F(\mathcal{O}(n)) = F(\varepsilon_{P'}(n_{P'}) g_{P'}(n)^* \Gamma) + O(|n - n_0|/N),$$

and thus by dividing  $P'$  into sufficiently small (but still dense) sub-products, we may assume that

$$|\mathbb{E}_{n \in P'} F(\varepsilon_{P'}(n_{P'}) g_{P'}(n)^* \Gamma)| \gg 1$$

for some  $n_{P'} \in P'$ , which by the total equidistribution of  $g_{P'}$  implies that

$$\left| \int_{G_{P'}/\Gamma_{P'}} F(\varepsilon_{P'}(n_{P'}) x) d\mu_{G_{P'}/\Gamma_{P'}} \right| \gg 1.$$

As  $F$  has vertical frequency  $\eta$ , this implies that  $\eta$  must annihilate  $G_{P', \geq d}$ , and so  $F$  is invariant with respect to the action of this group. By quotienting out by this

central group we may thus assume that  $G_{P', \geq d}$  is trivial, thus  $G_{P'}/\Gamma_{P'}$  now has degree  $< d$ . We can then write

$$\chi(n) = \tilde{F}(g_{P'}(n)^*\Gamma_{P'}, \frac{n}{5N} \bmod 1)$$

for all  $n \in [[N]]$ , where  $\tilde{F} : *(G_{P'}/\Gamma_{P'} \times \mathbb{T})$  is defined so that

$$\tilde{F}(x, \frac{n}{5N}) = F(\varepsilon_{P'}(n)x)$$

for  $n \in [N]$  and  $x \in *(G_{P'}/\Gamma_{P'})$ , and extended in a Lipschitz function to all of  $*(G_{P'}/\Gamma_{P'} \times \mathbb{T})$ . This represents  $\chi$  as a nilsequence of degree  $< d$  on  $P'$ . Using the conjugate nature of the various sequences  $g_P$  in Theorem D.5, we conclude that  $\chi$  can also be represented as a nilsequence of degree  $< d$  on all translates  $P' + h$  of  $P'$  also. On the other hand, since  $P'$  is dense in  $[[N]]^k$ , one can partition  $1 = \sum_{j=1}^J \psi_j$  on  $[[N]]^k$ , where  $J$  is bounded and the  $\psi_j$  are degree  $\leq 1$  nilsequences, each of which is supported on a translate  $P' + h_j$  of  $P'$ . This implies that  $\chi = \sum_{j=1}^J \psi_j \chi$ . As  $d \geq 2$ , the  $\psi_j$  have degree  $< d$ , and the claim now follows from Corollary E.2.  $\square$

We have the following useful consequence of Lemma E.11.

**Corollary E.12** (Extrapolation lemma). *Let  $k, d \in \mathbb{N}^+$  with  $d \geq 2$ , let  $\chi$  be a degree  $d$  nilcharacter on  $*\mathbb{Z}^k$  (with the degree filtration), and let  $N$  be an unbounded limit natural number. Let  $P_1, \dots, P_k$  be dense subprogressions of  $[[N]]$ , and let  $P := P_1 \times \dots \times P_k$ . Then the following are equivalent:*

- $\chi$  is  $< d$ -biased on  $[[N]]^k$ .
- $\chi$  is  $< d$ -biased on  $P$ .
- $[\chi]_{\Xi^d([[N]]^k)} = 0$ .
- $[\chi]_{\Xi^d(P)} = 0$ .

*Proof.* We trivially have that (iii) implies (iv). Since  $\chi$  correlates with itself, we see that (iii) implies (i) and (iv) implies (ii). Lemma E.11 gives that (i) or (ii) both imply (iii), and the claim follows.  $\square$

The Pontraguin dual  $\mathbb{T}$  of the integers  $\mathbb{Z}$  of course contains plenty of torsion. It turns out however that this torsion is a purely degree 1 phenomenon, and disappears in higher degree.

**Lemma E.13** (Torsion-free lemma). *Let  $k \in \mathbb{N}^+$ , let  $N$  be an unbounded integer, and let  $d \geq 2$  be standard. Then the abelian group  $\text{Symb}^d([[N]]^k)$  (with the degree filtration) is torsion-free.*

*Proof.* Our goal is to show that if  $q \geq 1$  is bounded and  $\chi$  is a degree  $\leq s$  nilcharacter such that  $\chi^{\otimes q}$  is equal to a degree  $< s$  nilsequence on  $[N]^k$ , then  $\chi$  is also equal to a degree  $< s$  nilsequence.

We modify the arguments used to prove Lemma E.11. We write  $\chi = F \circ \mathcal{O}$  where  $G/\Gamma$  is a degree  $\leq s$  nilmanifold,  $\mathcal{O} \in *\text{poly}(\mathbb{Z}^k \rightarrow G/\Gamma)$ , and  $F \in \text{Lip}^*(G/\Gamma)$  has a vertical frequency  $\eta$ , then we have

$$|\mathbb{E}_{n \in [N]^k} F(\mathcal{O}(n))^{\otimes q} F_0(\mathcal{O}_0(n))| \gg 1$$

for some degree  $< s$  nilmanifold  $G_0/\Gamma_0$ , some  $\mathcal{O}_0 \in *\text{poly}(\mathbb{Z}^k \rightarrow G_0/\Gamma_0)$ , and  $F_0 \in \text{Lip}^*(G_0/\Gamma_0)$ . Using Theorem D.5, we may thus find a product  $P = P_1 \times \dots \times P_k$  of progressions in  $[[N]]^k$  and a factorisation

$$(\mathcal{O}(n), \mathcal{O}_0(n)) = (\varepsilon_P(n)g_P(n)^*\Gamma, \varepsilon_{P,0}(n)g_{P,0}(n)^*\Gamma_0)$$



where  $\varepsilon_P \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G)$ ,  $\varepsilon_{P,0} \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow G_0)$  are bounded and Lipschitz on  $[[N]]^k$  with Lipschitz constant  $O(1/N)$ , and  $(g_P, g_{P,0}) \in {}^*\text{poly}(\mathbb{Z}^k \rightarrow \tilde{G}_P)$  is totally equidistributed in  $\tilde{G}_P/\tilde{\Gamma}_P$  for some rational subgroup  $\tilde{G}_P$  of  $G \times G_0$ . Shrinking  $P$  if necessary as in the proof of Lemma E.11, we may assume that

$$\left| \int_{\tilde{G}_P/\tilde{\Gamma}_P} F(\varepsilon_P(n_P)x)^{\otimes q} F_0(\varepsilon_{P,0}(n_P)x_0) d\mu_{\tilde{G}_P/\tilde{\Gamma}_P}(x, x_0) \right| \gg 1$$

for any  $n_P \in P$ . From the vertical character nature of  $F$ , this implies that  $\eta^q$  annihilates  $(\tilde{G}_P)_s$ . But  $\eta$  is a continuous homomorphism on the connected abelian Lie group  $(\tilde{G}_P)_s$ , and so  $\eta$  itself must also annihilate  $(\tilde{G}_P)_s$ . If we then quotient by this space, we can represent  $\chi$  by a degree  $< s$  nilsequence on  $P$ , and the claim now follows from Corollary E.12.  $\square$

#### APPENDIX F. A LINEARISATION RESULT FROM ADDITIVE COMBINATORICS

In this appendix, we record a lemma from additive combinatorics (essentially in [16] or [21], and in the spirit of Freiman's inverse sumset theorem) which asserts that functions from a large subset of  $[-N, N]$  to  $\mathbb{T}$  with a large amount of additive structure are essentially bracket-linear in nature.

**Lemma F.1** (Linearisation lemma). *Let  $\varepsilon > 0$  be a limit real, let  $N$  be a limit natural number, let  $H$  be a dense subset of  $[[N]]$ , let  $\alpha \in {}^*\mathbb{T}$  be a frequency, and let  $\xi_1, \xi_2, \xi_3, \xi_4 : H \rightarrow {}^*\mathbb{T}$  be limit functions such that*

$$\xi_1(h_1) + \xi_2(h_2) + \xi_3(h_3) + \xi_4(h_4) = \alpha + O(\varepsilon) \quad (\text{F.1})$$

*for many additive quadruples  $(h_1, h_2, h_3, h_4) \in H$ . Then there exists a standard  $k \geq 0$ , a frequency  $\delta \in {}^*\mathbb{T}$ , a dense subset  $H'$  of  $H$ , and a Freiman homomorphism  $\xi : H' \rightarrow {}^*\mathbb{T}$  of the form*

$$\xi(h) = \sum_{k=1}^K \{\alpha_k h\} \beta_k \pmod{1}$$

*for all  $h \in H'$  and some  $\alpha_k \in {}^*\mathbb{T}$  and  $\beta_k \in {}^*\mathbb{R}$  and some standard  $K$ , such that*

$$\xi_1(h) = \xi(h) + \delta + O(\varepsilon) \quad (\text{F.2})$$

*for many  $h \in H$ .*

*Proof.* We may replace  $\varepsilon$  by  $1/M$  for some limit integer  $M$ . By rounding each  $\xi_i(h)$  to the nearest multiple of  $1/M$ , we may assume that  $\xi_i(h)$  is a multiple of  $1/M$  for all  $h \in H$  and  $i = 1, 2, 3, 4$ . There are now only a bounded number of possibilities for the right-hand side  $\alpha + O(\varepsilon)$ , so by the pigeonhole principle (and by redefining  $\alpha$  if necessary) we may assume that

$$\xi_1(h_1) + \xi_2(h_2) + \xi_3(h_3) + \xi_4(h_4) = \alpha$$

for many additive quadruples  $(h_1, h_2, h_3, h_4)$  in  $H$ .

For each  $i = 1, 2, 3, 4$ , let  $\Gamma \subset {}^*\mathbb{Z} \times {}^*\mathbb{T}$  be the (limit) graph  $\Gamma_i := \{(h, \xi_i(h) \pmod{1}) : h \in H\}$ . Then by the preceding discussion, we see that  $(0, \alpha)$  has  $\gg N^3$  representations of the form  $\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$ , where  $\gamma_i \in \Gamma_i$  for  $i = 1, 2, 3, 4$ . On the other hand, from several applications of the Cauchy-Schwarz inequality, the number of

such quadruples is bounded by  $\prod_{i=1}^4 E(\Gamma_i)^{1/4}$ , where  $E(\Gamma_i)$  is the number of additive quadruples in  $\Gamma_i$  (i.e. the *additive energy* of  $\Gamma_i$ ). Since we have the trivial upper bound  $E(\Gamma_i) \ll N^3$  for all  $i$ , we conclude that

$$E(\Gamma_1) \gg N^3.$$

At this point we invoke some standard additive combinatorial machinery from [21] (see also [16, 51]). Applying the Balog-Szemerédi-Gowers lemma followed by the Plünnecke-Ruzsa inequalities exactly as in [21, Proposition 5.4], we can find a dense subset  $\Gamma'$  of  $\Gamma_1$  such that  $|9\Gamma' - 8\Gamma'| \ll N$ . Applying [21, Lemma 9.2], we can refine to a further dense subset  $\Gamma'' := \{(h, \xi(h) \bmod 1) : h \in H''\}$  such that  $4\Gamma'' - 4\Gamma''$  is a graph; thus there exists a Freiman homomorphism<sup>13</sup>  $\zeta : 2H'' - 2H'' \rightarrow \mathbb{T}$  such that

$$\xi_1(h_1) + \xi_1(h_2) - \xi_1(h_3) - \xi_1(h_4) = \zeta(h_1 + h_2 - h_3 - h_4) \quad (\text{F.3})$$

for all  $h_1, h_2, h_3, h_4 \in H''$ . By the Bogolyubov lemma (see [21, Lemma 6.3]),  $2H'' - 2H''$  contains a dense regular Bohr set  $B$  of bounded rank (see [21] for definitions; strictly speaking, one has to identify an interval such as  $[10N]$  with  $\mathbb{Z}/20N\mathbb{Z}$  in order to apply these tools, but this is not difficult to do). Arguing as<sup>14</sup> in [21, Proposition 10.8], we see that we may write

$$\zeta(h) = \sum_{j=1}^k \{\alpha_j h\} \beta_j \bmod 1$$

for  $h \in B$  for some standard  $k$  and frequencies  $\alpha_j, \beta_j$ . Applying (F.3) and the pigeonhole principle, we obtain the claim, except possibly for the claim that  $\xi$  is a Freiman homomorphism. But observe that if we restrict the fractional part of  $\{\alpha_j h\}$  to a sub-interval of  $I_0$  of length at most  $1/10$  (say) then we obtain the Freiman homomorphism property automatically; so the claim follows from one final application of the pigeonhole principle.  $\square$

## REFERENCES

- [1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn and D. Ron, *Testing low-degree polynomials over  $GF(2)$* , RANDOM-APPROX 2003, 188–199. Also: Testing Reed-Muller codes, IEEE Transactions on Information Theory 51 (2005), 4032–4039.
- [2] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica, **14** (1994), 263–268.
- [3] V. Bergelson and A. Leibman, *Distribution of values of bounded generalized polynomials*, Acta Mathematica **198** (2007), 155–230.
- [4] V. Bergelson, B. Host and B. Kra, *Multiple recurrence and nilsequences*, (with an appendix by I. Ruzsa), Invent. Math. **160**, no. 2 (2005), 261–303.
- [5] V. Bergelson, T. C. Tao and T. Ziegler, *An inverse theorem for uniformity seminorms associated with the action of  $\mathbb{F}_p^\infty$* , Geom. Funct. Anal. **19** (2010), no. 6, 1539–1596.
- [6] N. N. Bogolyubov, *Sur quelques propriétés arithmétiques des presque-périodes*, Ann. Chaire Math. Phys. Kiev **4** (1939), 185–194.
- [7] N. Bourbaki, *Groupes et algèbres de Lie*, vol I.
- [8] J. Bourgain, *On arithmetic progressions in sums of sets of integers*, in *A tribute to Paul Erdős*, CUP 1990.
- [9] ———, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.

<sup>13</sup>The notion of a Freiman homomorphism was defined in §3.

<sup>14</sup>This proposition involved a quadratic function on a Bohr set, rather than a linear one, but it is clear that the argument specialises to the linear case.

- [10] J. P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales*, C. R. Acad. Sci. Paris **306** (1988), 491–493.
- [11] N. Frantzikinakis, B. Host and B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, J. Reine Angew. Math. **611** (2007), 131–144.
- [12] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
- [13] H. Furstenberg, *Nonconventional ergodic averages*, The legacy of John von Neumann (Hempstead, NY, 1988), 43–56, Proc. Sympos. Pure Math., 50, Amer. Math. Soc., Providence, RI, 1990.
- [14] H. Furstenberg, B. Weiss, *A mean ergodic theorem for  $1/N \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$* . Convergence in ergodic theory and probability (Columbus, OH, 1993), 193–227, Ohio State Univ. Math. Res. Inst. Publ., 5 de Gruyter, Berlin, 1996.
- [15] K. Gödel, *Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory*, Proc. Nat. Acad. Sci. **24** (1938), 556–557.
- [16] W. T. Gowers, *A new proof of Szemerédi’s theorem for progressions of length four*, GAFA **8** (1998), no. 3, 529–551.
- [17] ———, *A new proof of Szemerédi’s theorem*, GAFA **11** (2001), 465–588.
- [18] B. J. Green, *Arithmetic progressions in sumsets*, GAFA **12** (2002) no. 3, 584–597.
- [19] ———, *Generalising the Hardy-Littlewood method for primes*, International Congress of Mathematicians. Vol. II, 373–399, Eur. Math. Soc., Zurich, 2006.
- [20] B. J. Green and T. C. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.
- [21] ———, *An inverse theorem for the Gowers  $U^3$ -norm, with applications*, Proc. Edinburgh Math. Soc. **51**, no. 1, 71–153.
- [22] ———, *Quadratic uniformity of the Möbius function*, Annales de l’Institut Fourier (Grenoble) **58** (2008), no. 6, 1863–1935.
- [23] ———, *Linear equations in primes*, Ann. Math **171** (2010), no. 3, 1753–1850.
- [24] ———, *The quantitative behaviour of polynomial orbits on nilmanifolds*, preprint available at arXiv:0709.3562.
- [25] ———, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, Contrib. Discrete Math. **4** (2009), no. 2, 1–36.
- [26] ———, *The Möbius function is strongly orthogonal to nilsequences*, preprint available at arXiv:0807.1736.
- [27] ———, *An arithmetic regularity lemma, associated counting lemma, and applications*, in An Irregular Mind (Szemerédi is 70), Bolyai Society Mathematical Studies **21**, 261–334.
- [28] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers  $U^4[N]$  norm*, Glasgow Math. J. **53** (2011), 1–50.
- [29] ———, *An inverse theorem for the Gowers  $U^{s+1}[N]$ -norm*, announcement available at arXiv:1006.0205.
- [30] I. J. Håland, *Uniform distribution of generalized polynomials*, Ph.D. Thesis, The Ohio State University, 1992.
- [31] ———, *Uniform distribution of generalized polynomials*, J. Number Theory **45** (1993), 327–366.
- [32] ———, *Uniform distribution of generalized polynomials of the product type*, Acta Arith. **67** (1994), 13–27.
- [33] I. J. Håland and D. E. Knuth, *Polynomials involving the floor function*, Math. Scand. **76** (1995), 194–200.
- [34] B. Host and B. Kra, *Convergence of Conze-Lesigne averages*, Erg. Th. Dyn. Sys. **21** (2001), 493–509.
- [35] ———, *Averaging along cubes*, Dynamical Systems and related topics, Eds. Brin, Hasselblatt, Pesin. Cambridge University Press, Cambridge (2004).
- [36] ———, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488.
- [37] ———, *Uniformity seminorms on  $\ell^\infty$  and applications*, J. Anal. Math. **108** (2009), 219–276.
- [38] B. Host, B. Kra, A. Maass, *Nilsequences and a structure theorem for topological dynamical systems*, Adv. Math. **224** (2010), no. 1, 103–129.
- [39] E. Hrushovski, *Totally categorical structures*, Trans. Amer. Math. Soc. **313** (1989), no. 1, 131–159.

- [40] R. Jin, *Freiman's Conjecture and Nonstandard Methods*, preprint.
- [41] T. Kaufman, S. Lovett, *Worst Case to Average Case Reductions for Polynomials*, preprint.
- [42] M. Lazard, *Sur les groupes nilpotents et les anneaux de Lie*, Ann. Sci. Ecole Norm. Sup. (3) **71** (1954), 101–190.
- [43] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206.
- [44] A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60.
- [45] ———, *Pointwise convergence of ergodic averages of polynomial sequences of translations on a nilmanifold*, Ergodic Theory and Dynamical Systems **25** (2005), no. 1, 201–213.
- [46] ———, *A canonical form and the distribution of values of generalised polynomials*, preprint.
- [47] E. Lesigne, *Équations fonctionnelles, couplages de produits gauches et théorèmes ergodiques pour mesures diagonales*, Bull. Soc. Math. France **121** (1993), no. 3, 315–351.
- [48] A. Mal'cev, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk SSSR, Ser. Mat. **13** (1949), 9–32.
- [49] W. Magnus, A. Karras, and D. Solitar, *Presentations of Groups in Terms of Generators and Relations*, Dover Publications, 2004.
- [50] K. Ribet, *On  $\ell$ -adic representations associated to modular forms*, Inventiones math. **28**(1975), 245–275.
- [51] A. Samorodnitsky, *Low-degree tests at large distances*, STOC 2007.
- [52] B. Szegedy, *Higher order Fourier analysis as an algebraic theory I*, preprint.
- [53] B. Szegedy, *Higher order Fourier analysis as an algebraic theory II*, preprint.
- [54] B. Szegedy, *Higher order Fourier analysis as an algebraic theory III*, preprint.
- [55] T. C. Tao, *Structure and Randomness*: pages from year one of a mathematical blog, American Mathematical Society, 2008.
- [56] T. C. Tao and T. Ziegler, *The inverse conjecture for the Gowers norms over finite fields via the correspondence principle*, Anal. PDE **3** (2010), no. 1, 1–20.
- [57] T. C. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, **105**. Cambridge University Press, Cambridge, 2006.
- [58] ———, *John-type theorems for generalized arithmetic progressions and iterated sumsets*, Adv. in Math. **219** (2008), 428?–449.
- [59] I. M. Vinogradov, *The method of Trigonometrical Sums in the Theory of Numbers*, translated from the Russian, revised and annotated by K.F. Roth and A. Davenport, Interscience, London, 1954.
- [60] T. Ziegler, *Universal Characteristic Factors and Furstenberg Averages*, J. Amer. Math. Soc. **20** (2007), 53–97.

CENTRE FOR MATHEMATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE CB3 0WA, ENGLAND

*E-mail address*: `b.j.green@dpms.cam.ac.uk`

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095, USA

*E-mail address*: `tao@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA, ISRAEL 32000

*E-mail address*: `tamarzr@tx.technion.ac.il`